Use of data to improve Seasonal-to-Interannual forecasts simulated by Intermediate Coupled Models.

by

Claire M. Perigaud (*), Christophe Cassou (*), Boris Dewitte (*), Lee-Lueng Fu (*),

and J. David Neelin (**)

(*) Ocean Science Element, Jet Propulsion Laboratory, California Institute of Technology

(**) Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California Los Angeles

Submitted to Mon. Wea. Rev.

March 21, 1999

Corresponding author: Claire Perigaud,
Jet Propulsion Laboratory, MS 300/323,
4800 Oak Grove Drive,
Pasadena, CA 91109.
email : cp@pacific.jpl.nasa.gov
Phone: (818) 354 82 03
Fax: (818) 393 6720

Abstract

This paper provides a detailed illustration that it can be much more beneficial for ENSO forecasting to use data to improve the model parameterizations rather than to modify the initial conditions to gain in consistency with the simulated coupled system. It is shown that persistence is not a reasonable benchmark for comparison to model forecasts, because sine functions fitted to observed Niño3 indices prior to forecasting have a correlation well above the one given by persistence. It is also argued that the statistics commonly used to judge the predictive skills of models easily lead to false positive results and unwarranted inferences. It is preferable to analyze the model behavior as time increases from the data-forced into the data-free context of a forecast experiment. Using sea level data in addition to wind to initialize the Cane and Zebiak model does not improve El Niño forecasts. This is because this model does not reproduce realistic sea level anomalies in a data-free coupled context. Nudging the observed wind to the model one to initialize the forecasts does not correct the model deficiencies and degrades the initial conditions with unrealistic sea level and wind anomalies. Using various oceanic and atmospheric data to change the model parameterizations allows to significantly improve the initial conditions and the forecasts up to 6 month-lead-time. This result is independent of the various initialization procedures used in this study. However, because of erroneous winds simulated by the atmospheric component in the eastern Pacific, cases with fast error growth are frequent. Replacing the atmospheric model by a statistical one results in more reliable predictions over 1980-1998. For lead-times up to one year, the forecasts predict well the observed anomalies between 1984 and 1993, including the sea level rises along the ITCZ after warm events and their following equatorward migration. This success is attributed to the consistency between the observed anomalies over this period and the mechanisms involved in maintaining the oscillatory behavior of the model, including the off-equatorial meridional wind anomalies.

coupled ocean-atmosphere models

1. Introduction

Despite a long history in predicting short-range climate variations and the availability of a wide variety of ENSO forecast models today (for a review, see Latif et al, 1998), the use of oceanic data to initialize forecasts is fairly recent. The present study started in December 1991, a few months prior to the launch of the TOPEX-Poseidon satellite. The objective was then to prepare the procedures to use the very accurate altimetric data delivered by this satellite for improving forecasting. Using sea level data to initialize a coupled model is indeed likely to have a long-lasting impact on forecasts, because these data monitor the vertically integrated density of the ocean which happens to have a much longer memory than the atmosphere. A few studies have thus illustrated the importance of the oceanic data assimilation on predictions (Kleeman et al, 1995; Ji and Leetma, 1997; Rosati et al, 1997; Fisher et al, 1997). In addition, coupled models are very sensitive to small changes in their initial conditions, so it is likely that the accuracy of the data used in the initialization process is a crucial issue.

This study uses the model presented in Zebiak and Cane (1987), CZ hereafter, and sea level estimates that are derived either from eXpendable BathyThermograph (XBT) since 1980 or from TOPEX-Poseidon since October 1992 (see Appendix A). The CZ code, the climatological files and the standard procedure to initialize the forecasts including the regular updates of the wind anomalies since 1964 are all provided by Dr. Zebiak from Lamont Doherty Earth Observatory (LDEO). Our original approach consisted in improving the initial conditions of the model without modifying the code itself. We first developed and tested various methods for initializing the model with sea level data. In particular we decomposed the dynamic height anomalies on their Kelvin and Rossby components as in Perigaud and Dewitte (1996) and nudged them into the baroclinic ocean over 1980-1995, or we applied the optimal Kalman filter developed in (Fukumori, 1995) for a baroclinic model similar to CZ to assimilate the TOPEX-Poseidon data over 1992-1995. Various procedures to initialize the mixed layer and atmosphere with the assimilated baroclinic fields, with the observed SST, and/or with the observed winds were also tested. Very little of this work has ever been reported (part of it is given here in Appendix A), because all these assimilation methods fail to improve the forecasts. Only one series produced by one of them is presented in this paper, for comparison with the CZ series. The latter are initialized using either the LDEO1 procedure described in Cane et al (1986), or the LDEO2 procedure described in Chen et al (1995a). The sole difference between the two procedures is the wind used to initialize the system: each time the baroclinic model is integrated in an initialization step increment, LDEO1 forces it with the observed wind anomalies only, whereas LDEO2 nudges the latter with the model wind anomalies.

The experiments presented in this paper correspond to model simulations that are either "forced" or "coupled". "Forced" simulations refer to those where the model is integrated in time with some data input. They include all the experiments for which observations have been inserted in the model, whether it is wind, sea level, SST only or a combination of oceanic and atmospheric data, whether observations are introduced into the model via direct forcing, nudging or optimal assimilation. "Coupled" simulations refer to those where no data is ever introduced during the experiment, the model evolving freely after some "forced" initial conditions. They include multi-decadal "coupled" simulations and forecasts. Models behave very differently whether they are integrated in a "forced" or in a "coupled" context, and recent efforts have been devoted to make the "forced" initial conditions more consistent with the "coupled" model to reduce the shock at the transition. Thus the LDEO2 procedure prepares "forced" initial conditions which are more consistent with the "coupled" mode than LDEO1. It significantly reduces the number of forecast cases for which the error grows (Xue et al, 1997). It delivers forecasts which are extremely good over the period 1980-1992 (Chen et al, 1995a, 1997). It fails to predict the 1997-98 El Niño event, but after assimilating tide gauge sea level data in the baroclinic model and nudging the results with the model winds like in LDEO2, Chen et al (1998) succeed in predicting the last event with forecasts initiated from March 1997. Part of the present paper investigates the reasons why differences in the initialization procedures lead the same CZ model to successes or failures in forecasting.

Although most of the sea level data used in our study do not have the accuracy of TOPEX/Poseidon, the authors are convinced that their preliminary results fail neither because of data errors nor because of inadequate methods of assimilation. One of the purposes of this paper is to provide evidence that failure to forecast with CZ is due to inadequate model parameterizations. Indeed model deficiencies do not originate from missing physics only. The parameterizations of the CZ model were estimated more than 10 years ago when there were so few observations available, whereas today various data sets provide information about the atmosphere or the ocean that can be used to revise them. The data used in this study have been presented either in Perigaud and Dewitte (1996) or in Perigaud et al (1998). They all have been extremely helpful to improve the anomalies simulated by the model when it is integrated in a "forced" mode (Dewitte and Perigaud, 1996) or in a "coupled" mode (Cassou and Perigaud, 1998), respectively refered to as DP96 and CP98 hereafter. The XBT data provide the oceanic temperature vertical structure used to derive a new parameterization of subsurface temperature, "Tsub", in the mixed layer model, while the ISCCP (International Satellite Cloud Cover Project) and AVHRR (Advanced Very High Resolution Radiometer) data provide a spatio-temporal relationship between SST and heating released by atmospheric convection, "Conv", introduced in the atmospheric model. Both "Tsub" and "Conv" parameterizations are described in DP96 together with their impact on the anomalies simulated in a forced context. The model with the revised parameterizations, Tsub.Conv, behaves very differently than CZ over multi-decadal "coupled" integrations away from the initial conditions (see CP98). To further improve the coupled simulations, the atmospheric component was replaced by a statistical one in a model referred to as Tsub.Astat (also presented in CP98). After analyzing the behaviors of these three ICMs during forced or multidecadal coupled integrations, the present study examines their forecasting skills, based on series of two-year long coupled experiments initiated every month from various "forced" conditions.

The paper is organized as follows. Section 2 discuss the criteria commonly used to determine predictive skills of models, for the CZ forecasts or for the predictions obtained by best fitting sine functions to SST observations. This is done to highlight the limits of the conventional statistics and to explain why alternative analyses are applied in the rest of the paper. Forecasts delivered by the CZ model are analyzed in Section 3 for various initialization procedures. Forecast series delivered by Tsub.Conv and by Tsub.Astat are analyzed in sections 4 and 5 respectively. The paper is concluded with a summary and perspectives in section 6.

2. CZ model or sine fits to observed SST: Statistical Performances.

The question addressed in this section is about how to determine if a model has some skill or not in predicting seasonal to interannual variations. Usually one compares the *a-posteriori* predicted SST Niño3 indices with the observed ones. For a given set of forecasts, the conventional approach consists in sorting out the indices by lead-times from 0 to 24 months, 0 referring to the initial conditions. Each lead-time defines a time series of forecast indices covering the period analyzed in retrospect over which correlation and error with observations are computed. Persistence, which consists in predicting that the initial anomalies will stay the same for the next two years, is the usual benchmark. This section is devoted to demonstrating that such an approach is misleading.

The CZ model is used here to deliver 312 forecasts every month between January 1970 and December 1995 initialized with the standard LDEO1 procedure, meaning that the ocean component started from rest in 1964 has been forced by the "detrended" FSU wind stress anomalies and that the latter has provided the SST which has then been used to drive the atmospheric model. Statistics in Figure 1ab have been computed between January 1972 and December 1995. The predictions initiated in 1970 for lead-times less than 24 months have been dropped out, so that statistics for each lead-time are represented by the same number of forecast cases (288). Statistics are also computed over two different epochs with 144 cases each (see Fig.1ghij), because as mentioned for example in Chen et al (1995a), results are known to significantly vary from one decade to the other.

Over 1972-1995, correlation is higher than the 95% level of confidence for all lead-times and CZ performs better than persistence past 6 months (Fig1a). The error (the rms. difference between the predicted and observed indices) is also smaller than for persistence past 12 months (Fig1b). Note that the error plotted in Figure 1b represents the full error as it includes the mean value of the difference between forecasts and observations. For CZ, this mean value is large and drifts as a function of lead-time: it reaches 1°C for a 6 month lead-time and is maximum with 1.2°C at 14 months. The issue of accounting for the mean and drift or not has been discussed for instance, in Syu and Neelin (1998). Typically real-time forecast series are presented in the literature relative to the mean values that the model has at this particular lead-time. How the model means drift with lead-time is certainly a complementary information which deserves more attention. Relative to these drifting means, errors are found to be much smaller. For CZ, they would be close to the observed level of variability (0.8°C), whereas they are larger than this level as soon as 5 months after the initialization in Figure 1b. Following the choice done in Chen et al (1995a, 1997), we choose to include the means in the errors presented here.

This section is devoted to questioning the validity of such statistics for determining skills in prediction. Because El Niño has a variability peak centered around 4 years, persistence exhibits a correlation with observations that has a minimum around 24 month lead-time. This minimum is negative. If El Niño corresponded to an index that varies like a pure 4 year sine function, persistence would have a correlation coming back up after 24 months to reach a value of 1.0 at 48 month lead-time. This explains why so many models seem to have their skill improving again for lead-times longer than one year (for example, see the review by Latif et al, 1998). In this review like in all plots of Figure 1, this apparent gain of skill corresponds to increases in correlation that are significantly above the 95% levels of confidence. Similarly, it is found that the correlation reached by persistence for a lead-time equal to 48 months is significantly positive: The observed Niño3 index series between 1970 and 1993 is correlated with the one between 1974 and 1997 by 0.29, well above the 95% level of confidence (0.12) or even the 99% level (0.15). Note also that the validity of these levels can be questioned as they are based on the assumption that the series are randomly distributed, whereas neither the observed nor the model signals have a random distribution (see more details at the end of this section). In any case, because the observed El Niño is not a chaotic signal like weather, it does not make sense to choose persistence as a benchmark. Rather than assuming that the anomalies are going to stay the same over the next two years, one can predict that the Niño3 index is going to oscillate with some amplitude, phase and period using sine functions. The performance of such sine functions is examined in the following.

Let us first focus on the sine period. It is commonly believed that one needs to have a coupled model which is able to simulate oscillations with a period centered around 4 years in order to be able to have some skill in forecasting. In order to show that this is a false statement, we optimally determine the period, amplitude and phase of the individual sine functions by rms. fitting each of the 312 two-year long data segments between January 1970 and December 1995. Results show that most of the data segments are best fitted with a period close to 2 years, not 4 years (Fig 1c). The same exercise is performed with the CZ standard forecasts over the same epoch and it is also found that most of the sine fits have an optimal period close to 2 years (Fig 1d). Of course, part of the reason why it is 2 years and not 4, is the duration of the data segments, but it is not the only reason. In addition, there are 250 cases out of 312 for the CZ forecasts which have an optimal period comprised between 21 and 27 months versus 172 cases only for the data, meaning that most of the CZ forecasts do actually "oscillate" with a period close to 2 years (the reader can verify this directly in the diagrams presented in Figure 2). Thus, the period of oscillations simulated by a coupled model over multi-decadal experiments is not necessarily identical to the period of oscillations simulated during forecasts. This is why models analyzed in CP98 that do not exhibit oscillations at a 4 year period, are nevertheless used in sections 4 and 5 of this paper.

As a consistency check, we now specify the period of the sine functions to either 2 or 4 years and optimize only the amplitude and the phase that best fit each two-year long data segment. In Figure 1ef, the sine fits are performed over the segments of data that cover the time of the forecast. Therefore, these fits cannot be considered as forecasts, but they are sorted out per lead-time in a similar format, in order to examine the correlation and errors with observations and compare the performance of a 2 year versus 4 year sine period. This figure shows that both perform equally well.

In Figure 1ghij, sine fits are performed over 4 year long segments of data prior to the forecasts. The optimal amplitude and phase are then used to predict the following 2 years. Statistics of the sines having a period fixed to 4 years are presented over either 1974-1985 (Fig.1gh) or 1984-1995 (Fig 1ij) because during these two epochs and the last one in particular, El Niño is not considered as being in a "regular 4 year-mode of oscillation" (see e.g. Torence and Compo, 1998). It is found that the sine fits have a predictive skill that appears similar or better than CZ for the two epochs. They also appear better than persistence past 5 or 6 month lead-time.

These results indicate that persistence is not an appropriate benchmark. But they also lead to issues more important than the performance of sine functions as a benchmark. The important point made here is that these statistics can lead to false positive inferences. Indeed it is striking that most of the models (Goswami and Shukla, 1991; Barnett et al, 1993; Kleeman, 1993; Ji et al, 1994; Kirtman and Zebiak, 1997; Rosati et al, 1997; Syu and Neelin, 1998; Schneider et al, 1998) give fairly similar skill scores in retrospective forecasting, although the modeling and the coupling assumptions as well as the individual forecasts/hindcasts differ so substantially. The levels of confidence usually associated to these statistics are based on the number of forecasts only, assuming that the errors between the model and the observed seasonal to interannual signals have a random Gaussian distribution. This is not the case for CZ (see section 3). It is not the case either for most of the Coupled General Circulation Models that usually undergo drifts in time at the transition between the forced and the coupled integration. Moreover, models predict signals which, relative to their mean and drift, often correspond to one oscillation over a few years like the observed signals. It is thus not correct to assume that the model-data differences are randomly distributed. So ranking the skills of models based on correlation values being larger than the levels of significance provided by the number of forecast initializations only, is inappropriate.

To illustrate how relying solely on these statistics can be deceptive, we now make forecasts issued from systems that nobody would trust, like sine functions that have some arbitrary or random characteristics. For example, instead of fitting the sine functions with the information from the previous years as in Fig1ghij, we match only the initial value to the observed Niño3 index. Then, in addition of the period and amplitude, we need to choose the sign of the initial trend in order to fully determine the parameters of the sine function. The plain line in Figure 1k corresponds to the arbitrary choices of a period of 3 years, an amplitude of 2°C and a positive sign. This figure shows that correlations then stay above 0.3 for all lead-times. Similarly, other series of sine forecasts verifying the equation:

$$\text{forecast}(t, k) = a(k)\ \sin(2\pi (t - t0(k) / \text{period}) + \text{phi}(k)),$$

matching the observed index at time $t = t0(k)$ at each month k between January 1980 and December 1994, were generated by testing various values for the period, the amplitude $a(k)$ and the sign of the trend at $t0(k)$. For example, the other curves in Fig. 1k have a period equal to 3 years and a random amplitude between 0 and 2°C uniformly distributed over the 180 cases k. They differ only by the sign of their initial derivatives, which is either systematically positive, or systematically negative, or randomly selected over the

180 cases. Although the cases with negative or random signs have a correlation that drops to negative values after about one year, all series reach correlations larger than +0.3 during the second year. Other sine periods between 2 and 4 years and maximal amplitudes between 1°C and 4°C were tested and for all, correlations at long lead-times always reach values well above the 95% level of confidence (0.14 for 180 cases). The fact that a random sign or a cooling trend tend to give lower correlations compared to the warming trends, is presumably due to the asymmetry of the observed series with peaks being narrower than troughs. But more importantly, we know by construction that none of these sine forecast systems can be considered as having skill, so the spread in correlation is an indication that one should not rely too closely on these statistics. No matter how good or bad the individual forecasts are, the fact that they are oscillating over a few years, permits them to appear as having considerable skill by this measure, with positive correlations well above the 95% level of confidence. This is all the more troubling as 95% of the individual forecasts differ by more than 1°C at any time for any lead-time greater than 5 months. Finally, we increased the number of cases to 320 (between January 1970 and December 1996) because statistics are usually considered as more reliable when calculated with a larger sample of data. Figure 1l presents the results for the same 4 types of sine functions and parameters as in Figure 1k. Correlations are generally slightly weaker, between 0.2 and 0.3 after 15 months, but they are still well above the 95% level of confidence (0.11 for 320 cases). Thus because of the low-frequency characteristics of ENSO, conventional statistics do not reliably discriminate between one prediction system and another.

The bottom line is that conventional statistics should be regarded with more caution, because the "ENSO red spectrum can easily lead to unwarranted and incorrect inferences" (Wunsch, 1999). These considerations explain why the results in the rest of this paper will not be presented in terms of these conventional statistics. Correlations and errors as a function of lead-time average out information about the model behavior during the time integration of the forecast experiments. One of the alternatives proposed in the literature and used hereafter is the "spaghetti" diagrams like those in Figure 2. Other alternatives need to be considered as well and will be discussed later. An advantage of the "spaghetti" diagrams is their ability to provide information on how individual forecasts behave during the "coupled" simulations in continuation with the "forced" ones.

3. Forecasts delivered by the CZ model with various initialization procedures.

This section presents an analysis of the forecasts delivered by the CZ model after three different initialization procedures. The forecasts initialized with the standard LDEO1 procedure are referred to as "CZ.STD", those initialized with sea level data to "CZ.SL" and those with the LDEO2 relatively more "coupled" initial conditions to "CZ.CPLIC". Most of the forecast series shown in this paper start in January 1980 because sea level data are not available before that date. Series of SST Niño3 indices are presented with a zoom between January 1981 and January 1993 in Figures 2 and 3. For clarity, forecasts with only 12 month long segments are plotted per diagram, the top diagrams corresponding to the first 12 months like the ones presented in Chen et al (1995a; 1997), and the bottom ones to the second year of prediction. Let us examine the predicted fields, starting with the SST.

a) SST forecasts

Figure 2 indicates that most of the CZ.STD individual forecasts undergo a strong oscillation during the two years, with about half of a cycle during the first year and the other half during the second year. This is consistent with the results presented in Section 2. For most of the cases, the period of sine functions that best fit the CZ forecasts is 2 years. The dashed line in Figure 2b that represents the time series obtained at a 12 month lead illustrates why sorting out forecasts per lead-time is misleading. This line should not be interpreted as a continuous time series, because any two consecutive points along this line come from different coupled experiments and have undergone variations in their past 12 months that are clearly distinct. The CZ model initialized with the LDEO1 procedure tends to systematically predict growing warm events with peaks up to 3°C or 4°C after a year, that then decay back to normal or slightly cold conditions. This overwhelming tendency to predict warm anomalies explains why the errors are so large when the forecast series are not corrected by their ensemble mean. The reasons why the CZ model simulates a warm bias in a "forced" or "coupled" context have been analyzed in Perigaud and Dewitte (1996) and in Perigaud et al (1998) respectively.

One of the series of forecasts performed by the CZ model after initialization with sea level data is presented in Figures 2cd. The details of the initialization procedure are given in Appendix A. Even though the model initialized with sea level data does not show evidence of improved skill, sea level data certainly have a strong impact on the individual predictions. CZ.SL forecasts differ from the standard ones by more than 2°C most of the time. Comparing the dashed lines in Fig.2a and 2c indicates that the impact of sea level data on the initial SST is not as big (except for the 1982-83 warming case, see potential explanations in Appendix A). Sea level data have slightly improved the initialization of the warm events in mid-1987 and late 1991, they have also slightly improved the cold event in 1988, but the initial wind is still far from reaching the strong easterlies observed at that time (see DP96).

Similar type of behavior is reproduced both in CZ.STD and CZ.SL. Indeed, CZ.SL can also be well fitted by 2 year sine waves. The series predicted at a lead-time equal to 12 months (dashed lines in Figures 2b and 2d) are quite different from each other, but they both have sharp jumps from month to month conversely to the initial conditions in SST or sea level. After 12 months of forecast, the model has completely lost the memory of the oceanic initialization. CZ.STD and CZ.SL suffer from severe deficiencies, in particular from strong shocks that take place at the transition between the "forced" and the "coupled" mode.

The LDEO2 procedure was developed to reduce such shocks. This procedure is quite efficient and delivers very good predictions between 1982 and 1992 as can be seen in the diagrams representative of the forecasts named CZ.CPLIC (Fig 3ab). The reader is invited to refer to Appendix B to be aware of the detailed differences between this series and the one published as LDEO2 in the literature. CZ.CPLIC forecasts agree remarkably well with the observed SST over 1980-1993, even for lead-times bigger than 12 months. These forecasts are by far the most realistic predictions of SST Niño3 index over 1982-1992 among all the ones published so far. So one may argue that the procedure used to initialize the CZ.SL forecasts is inadequate, because it does not assimilate the sea level data in a context which is consistent with the physics of the coupled model. This is indeed a very valid concern and this is why we have tested various methods of sea level data assimilation in a more "coupled" context than LDEO1 like LDEO2. Some of them are reported in Appendix A. This is also why the LDEO2 procedure will be tested with the other models presented in this paper (see section 4). However, all methods applied to CZ led to forecasts that are as poor as in CZ.STD or CZ.SL. Chen et al (1997) also found that the LDEO2 procedure does not work when the nudging is applied to the SST instead of the wind. These successes and failures are all the more intriguing as the LDEO2 initialization procedure does not correct errors that are intrinsic to the model itself.

It can be expected that the LDEO2 procedure introduces features in the initial conditions that have the characteristics of the deficiencies found in the "coupled" simulations (see Perigaud et al, 1998). The latter are briefly recalled here: 1) The zonal wind stress anomaly along the equator has its maximum located almost always at the same place in the central Pacific, about 20° to the East of the dateline, whereas the observed ones are mostly centered around the dateline. 2) The thermocline in the western Pacific is way more upwelled than in reality during and after a warm event. 3) The equatorial zonal wind at the eastern boundary of the model has an opposite sign to the one in the Central Pacific, with a very strong unrealistic amplitude. 4) The zonal wind anomalies beyond 9° of latitude in the eastern Pacific are three times stronger than the strongest signal ever observed there. 5) The latter are associated to unrealistically large off-equatorial thermocline displacements, with signs that do not always match observations. 6) The off-equatorial wind and thermocline anomalies play a key role in the oscillatory behavior of the coupled model. The first point is illustrated in Fig.1 of Chen et al (1995a). The other points are successively examined below.

b) Thermocline in the equatorial wave guide

Figures 3cd present thermocline depth anomalies averaged over NiñoW (130°E-170°E, 5°S-5°N). Compared to observations, the initial thermocline simulated by CZ.CPLIC is shallower by about 20 m during most of the 1983 year (Fig3c). Actually it is always shallower than the observed one. This is not the case in the standard initialization. The tendency for the thermocline to rise in the western Pacific more than in observations is characteristic of the coupled CZ behavior and explained in Perigaud et al (1998). This cold reservoir in the western Pacific happens here to compensate the tendency to predict overly warm SST. During the forecasts, the thermocline becomes even much shallower than ever, way shallower than observations in 1983 (Fig.3d). Note that when the SST index is very well predicted like for example in April 1983 for lead-times equal 12 months (Fig.3b), the thermocline in NiñoW is three times shallower than the observed one (Fig.3d). These results question the validity of forecasts based on the SST signal only. Other fields like the thermocline must be examined as well. Here, the thermocline exhibits higher-frequency signals than the SST. It is actually pretty unusual in coupled simulations to find fields that do not have the same spectrum of variability. This happens with CZ because of the 9 month mobile mode (Mantua and Battisti, 1995). Note that even in this case where the oscillatory behavior of the thermocline differs from the SST, statistics for the thermocline sorted out per lead-times appear as successful as the SST. Analyzing lags between thermocline, wind and SST signals and their full amplitude at a given time is more informative. Observations often exhibit a lead of the wind (and thermocline) relative to the SST during the warming phase, whereas the CZ coupled model does not, it even introduces a lag in the wind after the warm peak (see Perigaud et al, 1998). This lag and the contribution of the off-equatorial coupling (see below) explain why the shallowest peaks of the thermocline are reached several months after the observed ones. The rises of the thermocline and SST warm growths happen in 1982, 1986 and 1990-1991 earlier than in the observations or the ones driven by FSU winds. They are associated with the erroneous location of the coupled westerlies along the equator and the role of the off-equatorial coupling between the wind curl and the thermocline (see below).

Analyzing time series only is not sufficient either, because indices can be misleading when box averages cover regions where anomalies are not homogenous. The choice of the box locations is usually based on where the observed variations are the largest, but the signals simulated by the coupled model can have very different spatial patterns. The observed variability of the thermocline displacements is compared in Figure 4 to the initial conditions simulated by the CZ model for the various procedures. The simulations obtained with LDEO1 agree to some extent with observations (Fig.4ab). Not surprisingly, the variability is much better reproduced by CZ.SL (Fig 4c). Indeed the difference between the latter and data is smaller than 5 m everywhere within 5° of the equator over the whole period. By contrast, it is striking that the initial conditions of CZ.CPLIC (Fig 4d) show a minimum in the central equatorial Pacific where the observed variability peaks. This node of variability associated with the fixed location of the equatorial wind anomaly is one of the deficiencies of the model when it is integrated in a coupled context (see Perigaud et al, 1998).

The agreement between the CZ.SL predictions and observations holds for a short while, whereas CZ.CPLIC forecasts are far away from reality. This is illustrated in Figure 5abcd where the thermocline anomalies along the equator are presented for some events, based on observations or one month lead forecasts. Prior to the 1982-83 (Fig 5a) and 1991-92 events (Fig 5d), the LDEO2 thermocline is severely tilted with a cold reservoir in the western Pacific which is opposite to what is observed. Actually during the warm events themselves (Fig 5c) or for increasing lead-times (Fig. 5e), the thermocline keeps on becoming shallower in the western Pacific way beyond the observed level. After the 1983 event (Fig 5b), it is shallow by almost 10 meters all along the equator. The upwelling of the thermocline all along the equator is systematically found after warm events during long coupled simulations (see Perigaud et al, 1998). Note that it is shallow in the eastern Pacific while the SST is still warm. Such a situation has never been observed during warm events, but it systematically happens in CZ after warm peaks, because the model assumes a very little impact of the thermocline displacements on the subsurface temperature in case of shoaling, the SST being mostly influenced by the anomalous surface current convergence. By contrast, the observed thermocline is shallow only in the central Pacific, it is slightly downwelled in the western Pacific while the observed SST is already back to normal or even slightly cold during this period. The unrealistically cold reservoir initiated by the LDEO2 procedure after 1983 certainly contributes to reduce the number of growing warm events predicted by CZ.CPLIC (Fig 5f). Regarding the CZ.SL forecasts, the model enters in a growth mode of warm events very fast, even though the model initialization obtained with the observed sea level and wind is very good. As soon as 3 months after initialization, the equatorial band is dominated by the signals associated with the erroneous zonal wind anomalies along the equator (see below). The thermocline predicted by CZ.SL in Figure 5f is very typical of the period simulated after warm peaks, with shallow positions in the western Pacific and a strong slope reversal in the far eastern Pacific where the wind blows eastward.

c) Zonal wind along the equator

Conversely to a common belief that the atmosphere has little memory, its initialization plays a key role for the CZ model because one of the nonlinearities of the coupled system comes from the initial guess provided to the atmosphere component to solve the iterative moisture convergence feedback (see Zebiak, 1986). For amplitudes of

SST Niño3 index greater than 0.1, the feedback term is initialized with the wind convergence of the previous time step and thus, the CZ atmosphere is not a "slave" to the ocean state. This is a reason why the LDEO2 procedure can have a strong impact on the initial wind conditions, even though it induces small changes in the initial SST conditions. Figure 6 presents along the equator the zonal wind anomalies delivered at the initial conditions by the LDEO2 procedure together with two estimates derived from observations, namely FSU and Astat, the wind reconstructed after the Singular Value Decomposition between SST and wind observations (see CP98). Astat winds will also be used in sections 4 and 5 of the present paper. Using these two observed estimates is useful here as they contain different information, the Astat winds having less meso-scale high-frequency variability than the FSU winds, but also missing part of the low-frequency variability in the western Pacific. Wind anomalies presented in Figure 6 are averaged over a minimum of 6 months in order to compare low-frequency signals only.

The inter comparison of the results indicates that the simulated wind is very different from observations, while the two observed estimates agree quite well. Even several months prior to the beginning of the observed warm event growth in 1982, the model wind coupled to the model SST exhibits strong westerlies in the central Pacific (Fig 6a). Actually as early as January 1986 and October 1990 (not shown), the LDEO2 also simulates strong westerlies in the Central Pacific, whereas the observed growth has not started yet. During warm events, the LDEO2 procedure simulates a wind which is blowing maximum to the East in the Central Pacific (close to 160°E), but also strong easterlies at the eastern boundary (Fig. 6bdf). The latter systematically appear during the simulated warm events and play a key role in the model behavior (Perigaud et al, 1997). They are not found in observations. They also explain the thermocline slope reversal along the equator illustrated in Figure 5f. Systematically during cold events like in 1984 or 1988, the LDEO2 procedure simulates strong westerlies at the eastern boundary which are not realistic either (Fig.6ce). Note that the initial wind simulated by LDEO1 exhibits similarly large mismatch with data. Nevertheless it is worth analyzing the LDEO2 wind like in Fig.6 rather than the LDEO1 wind, because it is the wind consistent with the LDEO2 oceanic conditions and it partly explains the thermocline mismatch with data presented in Figure 5. One needs to examine the off-equatorial anomalies to fully understand it though.

d) Off-equatorial wind and thermocline

Because the off-equatorial anomalies were found to play a crucial role in the coupled model behavior (see Perigaud et al, 1998), the thermocline and wind initiated by LDEO2 are now examined away from the equator. More specifically, warm SST conditions are associated with westerlies along the equator and easterlies beyond 9° of latitude, the latter being more than three times stronger than the strongest anomaly ever observed there. Indeed this is precisely because the model wind is less trusted than along the equator that the LDEO2 procedure applies a nudging coefficient that decreases with latitude, giving less weight to the model wind and more to FSU outside the wave guide. Figure 7ab presents the zonal wind stress anomalies averaged between 8°N and 12°N during 5 months centered around the warm events. In 1983, the observed wind has weak westerlies in the eastern Pacific, whereas the LDEO2 wind is very strong with the opposite sign. At that time, the wind simulated by CZ.STD has the same wrong sign as LDEO2, with an intensity as large as 0.8 Dyn/cm^2 (see the thin plain lines in Figure 7a). Although the nudging factor applied beyond 9° of latitude gives a weaker weight to the model (45%) than to the FSU wind (55%), the LDEO2 wind in the eastern Pacific is dominated by the erroneous model wind and has the wrong sign. Note that because the atmospheric model is not linear and depends on the model wind estimated at the previous time steps, the LDEO2 wind is not accurately obtained by simply applying the FSU and the CZ.STD wind with their respective nudging ratios at a given time. Consistently with the coupled behavior of the model, the same scenario is reproduced for all warm events. Thus in 1992 although the observed wind shows strong westerlies, the LDEO2 procedure simulates strong easterlies (Fig.7b).

Associated with the strong westerlies blowing at that time along the equator, the off-equatorial easterlies create a very strong cyclonic curl (the meridional wind stress is relatively much weaker than the zonal one in CZ). By "quasi-Sverdrup" balance, the pair of cyclonic curls on both sides of the equator is associated with shallow positions of the thermocline in the western Pacific. Figures 7c and 7d present the thermocline positions averaged over 6°N and 10°N during six months after the observed warm peaks in 1983 and 1992 respectively. Like in the coupled experiments with the CZ model, the thermocline simulated by the LDEO2 procedure for both events is then tilted to very shallow positions in the western Pacific. This is not specific to the North. Actually the shallowest of all the thermocline forecasts is found in the Southwestern Pacific, consistently with the results found in Perigaud et al (1998). In the Northwestern Pacific, the observed and simulated thermocline do not even have the same signs. In fact the observed thermocline is deeper than normal over the whole domain in the North at that time. Note that the thermocline simulated by LDEO1 is downwelled like in reality. The fact that the latter is deeper in 1983 than the observed one is due to the baroclinic friction which is too weak, as demonstrated when the LDEO1 procedure is applied with a friction equal to 6 months instead of 30 months (dashed lines in Fig. 7cd). Because the CZ friction is so weak, the LDEO1 off-equatorial variability is, as can be seen in Fig4, overly large compared to observations, but note that it is only because the LDEO2 wind has less high-frequency energy compared to FSU, that the level of thermocline variability simulated by LDEO2 appears in Fig. 4abd in better agreement with reality than CZ.STD. Figures 7cd indicate that in fact, the sign of the CZ.CPLIC thermocline in the off-equatorial western Pacific does not even match the observed one. The LDEO2 procedure has considerably degraded the initial conditions for the off-equatorial wind and thermocline. It has build up a cold reservoir in the western Pacific that compensate for the warming tendency of the model during forecasts. One can now understand why nudging the SST instead of the wind is not successful in improving forecasts (Chen et al, 1997): Off-equatorial model SST do not differ much from observed anomalies whereas off-equatorial winds do a lot.

e) Initialization with SL data

Recently the LDEO2 procedure has been completed with an assimilation of sea level data in a procedure named LDEO3, which has delivered good statistics for Niño3 indices over 1975-1997 (Chen et al, 1998). In 1996-1997, the sea level data have contributed to reduce the cold SST bias that LDEO2 is erroneously producing (see their Figure 2). Prior to 1996, differences between the LDEO2 and LDEO3 initial SST are very small. As the LDEO3 nudging for the wind is the same as in LDEO2, erroneous features in the wind similar to the ones described above are likely to be reproduced by LDEO3, even for the 1997-1998 event which reproduces the observed warm SST Niño3 index and therefore strong off-equatorial easterlies and thermocline rise. The fact that assimilating sea level data has so little impact is explained by the choice of the LDEO3 nudging in the baroclinic model. Contrary to the nudging of wind that gives more weight to data than model when distance from the equator increases, the nudging of sea level gives less. The weight drastically decreases from 80% at the equator with an e-folding scale of 2° in latitude, meaning that at 5°N or 5°S and beyond, the LDEO3 procedure retains the model thermocline up to at least 99%. Therefore, the LDEO3 procedure does not allow sea level data assimilation to correct the erroneous thermocline anomalies that are associated with the erroneous model wind which is retained up to 55% at 5° and 45% at 7° and beyond like in LDEO2. Not surprisingly, the sea level data have very little impact on the LDEO3 forecasts between 1975 and 1996 (the latter are very similar to the LDEO2 ones). The off-equatorial thermocline initialized by LDEO3 in late 1997-1998 is expected to be unrealistic in the western Pacific. Although the LDEO2 and LDEO3 procedures have greatly reduced the shock at the initialization of forecasts, they have severely degraded the oceanic initial conditions in comparison with the CZ.SL or the LDEO1 ones with a more reasonable friction. These results support the need for improving at first the model, before increasing the consistency of the initial conditions with the coupled model.

Kleeman et al (1995) are able to improve forecasts by assimilating observed thermocline depth anomalies into their coupled model. Our results illustrate that conclusions relative to the use of data to improve forecasts strongly depend on the model. Indeed the thermocline in their model has a more important role on the SST which does not depend on currents, neither baroclinic nor Ekman, nor climatological nor anomalous currents. It also has a more important role on the wind as their atmosphere, conversely to CZ, is effectively a "slave" to the ocean state via SST. Bennett et al (1998) assimilate thermocline depth and wind data into a model which has similar physics to CZ, the ocean being the same and the atmosphere being the CZ model without the iterative convergence feedback term. They point out to some model data inconsistencies that suggest that this type of ICMs misses some physics. It is certain that a model with only one vertical baroclinic mode for the ocean and the atmosphere and a fixed mixed layer depth, lacks a lot of the physics that play an important role in El Niño. Nevertheless, however simple a model is, very different behaviors can be simulated for a given set of equations, depending on which coupled mode grows among those allowed by the parameterizations (Neelin and Jin, 1993).

In summary, two ways of using data to improve forecasts have been examined. One consists in using more data to improve the initial conditions while the other consists in reducing the impact of data in order to increase the consistency of the initial state with the model. Both lead to very limited success. Rather than using data for the initial conditions only, a third option consists in using them to improve the model itself. As proposed in DP96 and CP98, data provide information to revise the parameterizations and allow to reduce the model deficiencies in a "forced" context as well as during multi-decadal "coupled" simulations. Let us now examine the series of forecasts simulated by the ICMs with the revised parameterizations.

4. Forecasts delivered by Tsub.Conv

The Tsub.Conv model is the CZ model where the parameterization of the subsurface temperature is modified according to oceanic measurements and where the iterative convergence feedback is replaced by a parameterization of convective heating derived from observations (see DP96 and CP98). Conv is an atmosphere model that resembles the one developed in Kleeman (1991) where the wind stress anomaly is diagnostically derived from SST. As recommended in Perigaud and Dewitte (1996) and unless specified otherwise, the friction applied in the baroclinic model is of the order of 3 to 6 months so that the thermocline variability driven by observed winds beyond 5° of latitude is in fair agreement with the observed one (see Fig.7cd). For these friction values, the model does not exhibit an oscillatory behavior in any of the coupled experiments presented in CP98. Tsub.Conv is first initialized as in LDEO1, except that the procedure is not started from rest in 1964, but from the LDEO1 initial conditions of December 1979. The Tsub.Conv initial states obtained since 1980 are improved not only in SST, but also in thermocline and wind (see DP96). Various series of forecasts were generated to test the sensitivity to the parameters that have been investigated in CP98. The reader can refer to Table 1 for a list of the experiments presented here.

a) Characteristics of forecasts initialized as in LDEO1.

Figure 8 inter compares Tsub.Conv (1) time series of SST and thermocline at a 6 month lead with data and CZ.STD forecasts. Results show that the Tsub.Conv model is not biased towards warm SST in the East nor upwelled thermocline in the West. The reduction of this bias corresponds to one of the two major improvements achieved by Tsub.Conv in comparison with CZ in simulating realistic coupled anomalies (CP98).

The other expected improvement is relative to the wind which variability should be better located. The zonal wind variability over 1980-1993 is presented for observations or forecasts with a 6-month lead in Figure 9. The observed reference used here is the Astat winds (Fig.9a). It will be used again in this section and the reader can verify in Perigaud et al (1998) that there is little difference with the corresponding map derived from FSU. Forecasts are derived from either CZ.STD (Fig.9b), CZ.CPLIC (Fig.9c) or Tsub.Conv (Fig.9d). The latter has the wind pattern which agrees best with observations. The wind maximum is located to the west of the other forecasts, only 10° to the east of the dateline. Note that the weak amplitude of Tsub.Conv is fixable by modifying the weight of convective heating in the atmosphere (see CP98) and will be addressed later in this section.

Although the errors are much smaller for Tsub.Conv than for CZ at a 6 month lead, they are large. Actually the Tsub.Conv (1) series correspond to forecasts that decay to the "quasi-normal" state when the lead-time increases (Fig 10a). The spaghetti diagrams illustrate that Tsub.Conv (1) misses the growth of all warm and cold events and has obviously a very poor predictive skill. Note that the conventional statistics for this series (not shown) indicate an apparently good performance with errors much smaller than for CZ.STD.

Tests were then done with a weaker friction in the baroclinic component so that the coupled model exhibits an oscillatory regime (this is the case for decay times larger than 12 months). Various weights given to the convective heating relative to the local heating in the atmosphere were also tested because the weaker the weight, the bigger the intrusion of westerlies in the eastern Pacific and the stronger the warm events (see CP98). As anticipated, Tsub.Conv forecasts are found very sensitive to the choices of these coefficients. The Tsub.Conv (2) series presented in Fig.10b has a 24 month friction and a convective weight of 10%. Then the coupled model simulates strong oscillations like in 1983 or 1997, with westerlies extending east of the dateline close to the observed position during strong events and still west of CZ. Tsub.Conv(2) forecasts do predict the 1982-1983 and 1986-1987 warmings. They also succeed to predict the 1988-1989 cooling and reversal. These apparently successful predictions hold up to about one year lead-time. But the reader can see that Tsub.Conv (2) has a tendency to predict systematic warm growth like CZ.STD. This is consistent with the warm bias coupled mode found in CP98 when the weight of the atmospheric convection is decreased. In addition results from CP98 allow to anticipate that it is for wrong reasons that the forecast indices appear good in 1982-83 and 1986-89. Both CZ with a 30 month-friction and Tsub.Conv with a 24 or even a 12 month-friction, need the off-equatorial anomalies of wind and thermocline to sustain interannual oscillations. Between 5° and 15° of latitude, strong curls are coupled to steep zonal tilts of the thermocline with respectively upwelled/downwelled waters during and after warm/cold events, the anomalies in the North Pacific not even having the right signs in comparison with data. The reader can see in Figure 9 erroneous wind patterns simulated by Tsub.Conv(1) in the off-equatorial eastern Pacific with some resemblance to CZ. During the 1982-83 event for example, we found that the one month lead forecasts delivered by Tsub.Conv (2) have erroneous off-equatorial wind stress and baroclinic anomalies as early as July 1982.

b) Forecasts initialized with sea level data.

The fact that the prediction of the Niño3 index depends on the off-equatorial ocean strengthens the need for a good ocean initialization, in particular away from the equator. Tsub.Conv (2.SL) is obtained with the same parameterization as Tsub.Conv (2), but has been initialized with sea level data. The procedure corresponds to the second method described in Appendix A that provides a better initialization of the baroclinic ocean beyond 5° of latitude in comparison with the first method used in CZ.SL. The impact of the sea level data on the initial SST anomalies is little (Fig. 10c), except in year 1982 and late 1989 for reasons given in the Appendix. Overall, the use of sea level data slightly reduces the initial error with the observed SST. The impact is small on forecasts, except those initiated in 1982 or late 1989. Tsub.Conv (2.SL) predictions resemble a lot the previous ones. The sea level data have a much smaller impact on Tsub.Conv (2) than on CZ (compare CZ.STD and CZ.SL). This is because the initial conditions, in particular the wind initialized by Tsub.Conv(2) or Tsub.Conv(2.SL) resemble a lot more FSU winds than CZ.STD or CZ.SL. So on the one hand, Tsub.Conv initial conditions are more realistic than CZ whether sea level data are used for initialization or not. On the other hand, Tsub.Conv has an unstable mode where coupled errors grow fast and make the system forget its good initialization soon.

Indeed the model has a tendency to grow warm events in Tsub.Conv(1) as well as Tsub.Conv(2). This unstable mode is present, no matter what friction nor convection coefficient is used. This is particularly striking during interevents when there is a discontinuity in the rate of change at the switch between the "forced" and the "coupled" simulations. The most important step to achieve is to reduce the growth of the unstable coupled mode in such cases. The mode developed by Tsub.Conv corresponds to warm SST anomalies associated to unrealistically large equatorial easterlies at the eastern boundary of the basin and has some resemblance with the one developed by CZ (see Fig 6 and 9).

c) Forecasts initialized with a more "coupled" procedure.

In order to reduce the initial shock undergone by forecasts, various initialization procedures were tested. The two presented now, Tsub.Conv (3) and Tsub.Conv (3.CPLIC) have been obtained with a 6 month friction and with the atmospheric convection weight fixed to the value of Tsub.Conv (1), because these parameters are consistent with observations in a "forced" mode and because whether the coupled model has an oscillatory regime or not, is not a necessary condition to deliver good forecasts (see next section).

Effort was put on increasing the consistency between the observed and the model wind as much as possible. Instead of using FSU data to force the ocean component, the wind comes from the Astat model presented in CP98 and shown in Figures 6 and 9 of this paper. In addition here, because the second SVD wind mode of the Astat model has its maximum amplitude in the eastern equatorial Pacific and favors the coupled mode growth (see CP98), only the first SVD mode is now retained to initialize the model. Tsub.Conv(3) and Tsub.Conv (3.CPLIC) forecasts are initialized following the LDEO1 and LDEO2 procedures respectively, but with this wind instead of FSU.

The Tsub.Conv (3) forecasts are presented in Figure 10d. By inter comparing the dashed line with the ones in Figure 10ab, it is shown that forcing the model with the Astat wind has some beneficial impact on the initial SST. In particular it removes the erroneous feature in 1989 (this is also the case when all the SVD modes are retained to force the model- see section 5) as well as the overly warm second warm peak in 1983 (this is the case when only the first SVD mode is retained- see Section 5). It is verified that the initial wind of Tsub.Conv (3) agrees very well with the Astat wind in the 10°S-10°N band (not shown). However different the initial conditions are, it is striking that most of the Tsub.Conv(3) forecasts resemble Tsub.Conv (1). Initial shocks are still big, forecasts are still bad. The model has a tendency to predict decaying warm events when it should predict growing warm events like in 1982, 1986 and 1992, growing warm events when it should predict the decays of warm events like in 1983 and 1992 or the growth of the cold event like in 1988. This suggests that the high-frequency fine-scale features of the FSU wind anomalies that could *a priori* be a source of error growth are not. Neither is the second SVD wind mode.

Finally, let us examine the Tsub.Conv (3.CPLIC) forecasts (Fig.10e). It is found that the LDEO2 procedure does not affect much at all the initial conditions nor the forecasts. It is quite remarkable that the impact on Tsub.Conv forecasts is so little, whereas it is so big on CZ. The initial shock is certainly somewhat reduced when the LDEO2 procedure is applied to Tsub.Conv, but the advantage of this reduction is small and lasts no more than 3 months (compare with Fig.10d). It is clear that the LDEO2 procedure can significantly affect the initial conditions only if the model wind is drastically different from reality, which is not the case for the Tsub.Conv model.

In summary, the use of data to revise the CZ model into the Tsub.Conv model has considerably reduced the errors in forecasting the SST Niño3 index (compare Fig 2a with Fig 10a or Fig 10d). Compared to CZ, the initial conditions and the forecasts during the first 6 months are much closer to observed thermocline, wind and SST anomalies all over the Pacific within 15° of the equator. Tsub.Conv forecasts are not much sensitive to changes in the initial conditions because the coupled wind and ocean resemble much more the observed ones than those simulated by CZ. Beyond a lead-time of 6 months, the model predicts a decay to the normal state when the friction decay time is 3 months. This is not the case with weaker frictions, but whatever its value or that of the atmospheric convective weight, Tsub.Conv skill is limited because of its erroneous mode of growth located in the eastern Pacific. Although the winds may not be the only model deficiency in the eastern Pacific, replacing the Conv by the Astat atmosphere is worth the investigation, because the coupled mechanisms developed by Tsub.Astat are quite different (see CP98).

5. Forecasts delivered by Tsub.Astat

It is found in CP98 that the Tsub.Astat model is much less sensitive than Tsub.Conv to parameter changes for the friction or the drag. It is also explained that the simulated wind stress and baroclinic anomalies in the North have opposite signs to the ones simulated by Tsub.Conv during and after warm peaks, in agreement with observations. Because the study is now focused on predictions, the data covering the period which is going to be predicted, should not be taken into account in the statistical atmosphere. It was clearly identified in CP98 that the second SVD mode plays a key role in the oscillatory regime. Decomposed over the 1970-1995 period, the amplitude of the second SVD wind mode reaches its maximum at the 1982-83 event. When the SVD decomposition is performed over 1970-1980 only, Tsub.Astat does not reproduce oscillatory regimes, whereas it does with the decomposition over 1970-1983. So rather than updating the SVD modes as time advances before each new forecasting year as is done in Syu and Neelin (1998), it is decided to fix them to those obtained over 1970-1983. We verified that the coupled behavior is similar to the one presented in CP98.

In the rest of the paper, the friction is fixed to 6 months. Like for Tsub.Conv, Tsub.Astat then does not oscillate during decade-long coupled experiments, it simulates a quasi-steady state around either warm, or cold or quasi-normal conditions. This choice is nevertheless justified because it corresponds to the one best fitting observations with the forced simulations, and results from the last section indicate that failure in predicting is not necessarily due to the non-oscillatory behavior of the model. In order to directly compare predictions that are initialized with the exact same oceanic conditions, the Tsub.Astat (1) forecasts are initialized by forcing the ocean with the first SVD wind mode as in Tsub.Conv (3). Tsub.Astat (1) and Tsub.Conv (3) forecasts are quite different (Compare Fig. 11a and Fig. 10d). Although there are still some cases during the decay of the 1983, 1987 and 1992 events that grow back warm again, many Tsub.Astat forecasts predict the reversals and decays with some skill. In addition, Tsub.Astat (1) does predict the warm growths past July 1982, past April 1986 and past May 1992, it does predict the cooling and reversal past March 1988. All these forecasts are improved compared to Tsub.Conv (3).

Tsub.Astat does not have the same error mode leading to warm growth in the eastern Pacific as Tsub.Conv. There is no reason for Tsub.Astat to be initialized by the first SVD wind mode only other than the above comparison. Indeed when it is initialized with the FSU wind (not shown), the model does not predict a warm growth after 1983 like in Tsub.Conv (1 or 2), it predicts a decay as in reality. The choice is then made to present the series of forecasts Tsub.Astat (2) obtained when Tsub.Astat is initialized by the wind reconstructed with all the seven SVD modes used in Astat as in CP98. Several series were generated with the various parameters tested in CP98. Consistently with the changes explained in this reference, forecast series are slightly modified. Two of them are presented here, the reader can refer to Table 2 for complete identification. As the quality of this series is considerably improved compared to all the previous ones, Tsub.Astat (2) forecasts are presented up to December 1997 in Figure 11b. The introduction of higher SVD modes, especially the second one, in the initial conditions contributes to "boost" the reversals and decays more clearly than in Tsub.Astat (1). Similar success is obtained with FSU indeed. This is a big difference with Tsub.Conv that simulates erroneous warm growth and overly warm initial conditions with FSU or the full SVD winds. Most of the individual Tsub.Astat (2) forecasts agree with data to some extent. The decays of the warm events in 1983 and 1987 with the continuation of the latter into the cold event in 1988, are captured as well as the warming mild trends between 1989 and 1994. Note that Tsub.Astat produces erroneous warm anomalies during late 1989 and underestimates the growing warm events of 1983 and 1997. Similar errors are found in most of the other published forecasts (Rosati et al, 1997; Kirtman and Zebiak, 1997; Syu and Neelin , 1998).

Because it is possible with Tsub.Astat to simulate stronger warm events by slightly changing the value of parameters while keeping realistic amplitude patterns and behavior simulated in the tropical Pacific within 15° of the equator (see CP98), several experiments are performed to test if the strength of the 1983 and 1997 events could be recovered without degrading the rest of the forecasting. Such a series with an increased coupling coefficient (Tsub.Astat (3), see Table 2) exhibit similar behavior and skill to Tsub.Astat (2) (Figure 11c). The predicted growths of the 1982 and 1997 warm events agree even better with observations, but they are still weaker and the model still fails to predict these events in forecasts started prior to May 1982 or prior to March 1997. Tsub.Astat (3) is no less skilled than the best published forecasts, the failure in predicting these events more than 6 months in advance deserves special attention (paper in preparation). At this point, it is rather necessary to examine and explain the good performance of Tsub.Astat in forecasting the Niño3 index for the 1984 to 1993 period. Validation with data is performed below for the wind, the SST and the thermocline anomalies over this period.

a) SST forecasts with one year lead.

In the next three paragraphs, the one year-lead Tsub.Astat (3) forecasts are inter compared with observations and with the CZ.CPLIC one year-lead forecasts during the 1987, 1988 and 1992 warm or cold events. The SST fields are presented in Figure 12. Tsub.Astat warm predictions in 1987 and in 1992 have an off-shore and a coastal maxima of about 1.4°C, whereas CZ.CPLIC predicts only coastal anomalies with a very large signal close to 2.6°C in 1987. These patterns and amplitude correspond to the features found in long coupled simulations for these two models (see CP98). The observed SST anomalies do not have a coastal maximum in 1987, but they exhibit a two-peak pattern in 1992 in very good agreement with Tsub.Astat. The cold event in Tsub.Astat is too weak and too widely spread in latitude on both sides of the equator, but it is central as in observations and not coastal as in CZ. Note that these maps illustrate that indices such as the SST Niño3 index can be misleading because of averaging effects.

b) Zonal wind stress forecasts with one year lead.

The corresponding observed and predicted zonal wind stress anomalies are presented in Figure 13. The westerlies in 1987 are predicted with a fairly good location and amplitude for Tsub.Astat, whereas CZ.CPLIC is much too strong and its maximum is located 20° to the east of the dateline. Similar success in forecasting the amplitude and patterns of the wind is achieved by Tsub.Astat in 1992. More importantly, CZ.CPLIC predicts very strong easterlies in the eastern Pacific, along the equator as well as off-equator. These are not found in observations. During the cold event, the wind pattern is really well predicted by Tsub.Astat. By contrast, the easterlies predicted by CZ.CPLIC are located 60° to the East of the dateline and thus happen to have a sign opposite to the one observed there.

c) Meridional wind stress forecasts with one year lead.

The meridional wind stress anomalies must also be validated as they play a strong role in maintaining the oscillatory regime via the off-equatorial thermocline and coupled adjustments. It is striking that the meridional component predicted by CZ.CPLIC is much weaker than the zonal one, whereas for data and Tsub.Astat, it reaches values that are larger in the North and in the Southwest (Fig. 14). Patterns and amplitudes related to the ITCZ and SPCZ displacements are remarkably well predicted by Tsub.Astat for both warm and cold events. By contrast, CZ.CPLIC winds are badly located in the eastern Pacific.

d) Meridional displacements of thermocline forecasts

Tsub.Astat (3) thermocline forecasts are now inter compared with observations and with CZ.CPLIC as a function of time and latitude. As explained in CP98, the oscillatory behaviors of both CZ and Tsub.Astat are sensitive to the charge and discharge of the oceanic heat content outside the equatorial band. This quantity can be monitored by the zonal average of the thermocline between the western and the eastern boundaries of the Pacific basin. Because El Niño is a low-frequency signal, this quantity is coupled to the wind stress curl anomalies via the quasi-Sverdrup balance rather than via Ekman pumping. The sign of the curl in the 5°N to 15°N band happens to be given by the meridional wind anomalies for Tsub.Astat, and by the zonal ones for CZ. Thus warm peaks simulated in long coupled runs are followed by a charge for Tsub.Astat and by a discharge for CZ of the oceanic heat content in the North. Let us examine if forecasts reproduce these scenarii and how they compare with observations.

It is striking that the observed 1984 to 1992 period is dominated by two features of thermocline anomalies coming from the North Pacific propagating southward and crossing the equator (Fig. 15a). The corresponding Hovmoeller diagrams were plotted for Tsub.Astat and CZ.CPLIC, either at the initial conditions or at the various lead-times up to 12 months. They all present the patterns of interest here. For simplicity, the average of the first 12 month lead forecasts is presented here. Tsub.Astat (Fig 15b) predicts reasonably well the observed features, except for the negative thermocline anomalies North of 5°N after 1994. The understanding of the deficiencies after 1994 needs further investigation. Between the 1983 and the 1987 warm peaks, the forecasts reproduce features that are consistent with the ones simulated in long-coupled experiments, meaning a charge in the North and a discharge in the South for Tsub.Astat (Fig. 15b) and a discharge on both sides of the equator for CZ.CPLIC (Fig. 15c). The forecast signals agree fairly well with the observed ones for Tsub.Astat, whereas they do not for CZ.CPLIC. Indeed the observed state of the ocean between 1983 and 1993 is compatible with the oscillations that the coupled Tsub.Astat model can simulate. Knowing the importance of the off-equatorial charge and discharge in the coupled oscillatory behavior, this state certainly contributes to the success of the Tsub.Astat model in forecasting the 1983 to 1993 period.

6. Discussion and perspectives

This paper uses several configurations of the CZ model to examine how data can be used to improve forecast skill. In particular, this model system is used to examine the impact of changing initialization schemes compared to using the data for improvement of model parameterizations and more complete validation. In the process of this examination, a number of questions were raised regarding the usefulness of the common measures of model predictive skill.

As a benchmark for predictive skill, persistence seems an unreasonably easy target to compare to. Sine functions with periods of 2 or 4 years fitted to data prior to forecasting are suitably simple and give better predictions. The period chosen for the sine fit is not as critical an issue as previously thought. Another surprise noted here is that models can oscillate with a period during multidecadal coupled experiments that is very different from the behavior developed during two-year long forecast series. The most important message is that the conventional statistics do not accurately represent the predictive skills of models. Seasonal-to-interannual climate variations are not randomly distributed and therefore the confidence levels usually given to measure the accuracy of such statistics are underestimated. Initialized to the observed index, sine forecasts with random amplitude appear skilled with correlations ranging between 0.2 and 0.4, independently of the number of forecasts generated. Sine fits cannot be used as benchmarks because their determinations cannot be unique. In addition to this uncertainty, statistics often hide biases or drifts that models undergo. Our experience indicates that paying attention to the latter is useful to understand the model behavior during forecasts.

Although it is usually thought that confidence in statistics is gained by extending the period of analysis, we have purposely limited our study to a relatively short duration in order to avoid general statements and rather highlight the uniqueness of each El Niño event. Results also suggest that models usually have some skill in predicting the reversals and decays of warm or cold events, whereas forecasting quasi-normal conditions after big events and during the many years of inter-events, is very challenging. It is similarly difficult to predict a big event one year in advance when the Niño3 SST index is still close to zero. So rather than sorting statistics per lead-time, one might focus on categories of initial conditions.

In addition, sorting out forecasts per lead-time renders the understanding of what the model predicts extremely difficult, because two consecutive points of a time series at a given lead-time, have different origins and past histories. It is thus impossible with statistics to assess the successes and failures of the physics simulated by the model when switching from the "forced" to the "coupled" context. Spaghetti diagrams are more appropriate for addressing this issue. Nevertheless they are not sufficient either. It is useful to analyze forecasts in terms of spatial patterns as well, since most indices rely on arbitrary limits defining regions where problems can be hidden by averaging. In addition, relying on the SST fields only, limits the understanding of the coupled processes that the model simulates. Compared to the other fields, SST is the one which has the most confined variability in space. Several examples have been given where Niño3 SST indices are reasonably well predicted, whereas thermocline in the western Pacific or winds in the eastern Pacific are not. Rather than proposing one alternate statistical approach to determine how good or bad model skills are, this study appeals for more diverse and accurate analyses of predictions over a relatively limited number of events. Finally, note that reaching skill at a 12 month lead-time is already quite a challenging performance. This is why very little attention is paid here to the second year of forecasts. More important than statistical performance, understanding the mechanisms at work during the forecasts is the key step in progress. Actually, the first year of forecasting is the one which deserves the most attention because of its unique situation at the transition with the initial conditions.

One of the objectives of this study is to demonstrate that data are useful in making progress in ENSO forecasting. Part of this paper examines how data can be used in the initialization procedures rather than in the improvements of the model itself. An initialization which gains in model consistency like the LDEO2 procedure, can very efficiently reduce the number of forecasts for which the errors grow. This is true when the model integrated in a "forced" context is relatively far from reality like CZ. Consistency of the initial conditions with the coupled model is certainly an issue worth to be addressed. When examined in terms of sea level and wind, the LDEO2 procedure delivers initial conditions that disagree with data beyond reasonable estimates of data uncertainty. More specifically, the wind anomalies beyond 5° of latitude are very strong with signs opposite to the observed ones, and are associated with very shallow thermoclines in the western Pacific that are not observed. The LDEO2 procedure actually consists in using relatively "less data" than the LDEO1 procedure. The LDEO3 assimilation of sea level data applied in addition to the LDEO2 wind nudging does not correct the thermocline nor wind errors because the model thermocline is given a much stronger weight than the data everywhere except within 2° of the equator. We regard these as cases where the procedures or the model should be revised. Using more data than in LDEO1 to improve the initial conditions, specifically sea level in addition to wind, does not work either. It is found useful to understand the mechanisms at work in multidecadal coupled simulations, because even if the period of oscillations is different during forecasts, the mechanisms of growth are the same. The CZ model has a tendency to predict warm events, associated with overly strong wind cyclonic curls in the off-equatorial eastern Pacific and an overly shallow tropical thermocline. If data are used to initialize forecasts only, model deficiencies are not corrected. Another option is to use data to reduce deficiencies of the model itself and thus make the physics simulated by the models more realistic.

The impact of improving model physics to match observations is illustrated here on forecasts delivered by ICMs. A number of caveats rise due to the simplified physics. For instance, the baroclinic model has only one vertical mode, so the choice of the time decay for the friction is important and the vertical structure of the ocean is not fully represented. The mixed layer has a constant depth and surface currents as well as their convergence may contain large errors. Although not mentioned here but very much like the coupled simulations in CP98, forecasts are highly sensitive to the values of anomalous and climatological upwelling. Keeping a simple modeling context, it is worth adding vertical modes in the baroclinic ocean like in (Chen et al, 1995b; Dewitte, 1998) and revising the formulation of the Ekman and mixed layer components like in (Boulanger, 1999). The Gill (1980) type of atmosphere, whether it is in the CZ or Conv version, has a serious flaw in generating spurious winds in the eastern Pacific. This appears not to be fixable by improving parameterizations while keeping this simplification for the atmospheric dynamics. For coupling to simple ocean models, alternatives include intermediate models (Wang and Li, 1993; Neelin and Zeng, 1998) or AGCMs (e.g., Kirtman and Zebiak, 1997). In the present study, the atmospheric model is replaced by a statistical approach. Whenever a coupled model has some statistics built into it, it is common to hear that the limited success in forecasting is primarily due to the statistical component of the model. Results in this paper highlight that this is not the case. The dynamic model itself needs to be improved. More complete physics is certainly the long-term future objective needed for climate predictions. Nevertheless, data can be used meanwhile for any model to improve its skill in simulating the reality. Used to revise the parameterization of CZ, data allowed to reduce errors when the model is either forced by some realistic conditions or free to evolve. It was not *a priori* certain that they would reduce error growth in coupled forecast mode.

The Tsub.Conv model simulates initial conditions and forecasts during the first 6 months that are much more realistic than the CZ model. However, this model still has a tendency to forecast warm conditions and erroneous winds in the eastern equatorial Pacific. Note that CP98 found that the long-term behavior of Tsub.Conv resembles more Tsub.Astat than CZ, whereas this study shows that forecasts delivered by Tsub.Conv resemble more CZ. The main difference between CZ and Tsub.Conv in terms of forecasting is that the Tsub.Conv initial conditions agree much better with observations, and changing the wind used in the initialization procedure, either by removing some of its energy in the eastern Pacific or by nudging it to the model wind following LDEO2, does not greatly affect individual forecasts.

The Tsub.Astat model is the most reliable prediction system of all the ones presented here. Forecasts do not depend as much on friction nor coupling coefficients. With a lead-time up to one year, the model performs very well in predicting the observed anomalies of SST, sea level, zonal and meridional winds, all over the Pacific between 15°S and 15°N for the period between 1983 and 1992. It predicts the oceanic recharge in the North coupled to the southward migration of the ITCZ that are observed after the 1983 and the 1987 warm events. Note that if the meridional wind stress anomaly is maintained to zero during a time integration of a forecast experiment, the model fails to predict the coming events. So it appears necessary to simulate realistic processes outside the equatorial band to have some skill in forecasting. This does not guarantee that Tsub.Astat predictions over this period give realistic anomalies for the right reasons. In particular, Tsub.Astat simulates surface current equatorial convergence anomalies that need validation. In addition, the success of a model in predicting a particular El Niño event does not guarantee its success for other events. Tsub.Astat fails to predict the warm trend prior to May 1982 or March 1997, whereas it succeeds in predicting the 1986-87, 1992 warm events and the 1988 cold events. Sea level data can be used to improve the initial conditions. In particular, TOPEX/Poseidon data are very efficient at improving the 5° to 15° bands of latitude prior to March 1997, but the reader can anticipate that this will not be sufficient to improve the forecasts (paper in preparation).

The priority in order to make progress in forecasting is to understand why a model is successful or not. The first step towards this understanding is to identify the model deficiencies either in a "forced" or in "coupled" mode and to reduce them. Because data errors are often much smaller than model errors, consistency of the initial conditions with the coupled model does not seem the priority today. This paper provides a detailed illustration of how data used to improve model parameterizations, can be very useful to improve predictions. This supports appeals from the community for the continuation of our monitoring system. In particular, TOPEX-Poseidon is needed to monitor the large scale low-frequency sea level variability associated with the charge/discharge of the coupled system. The appendix of this paper provides several examples where a change smaller than 1cm in the initial conditions has a strong impact on forecasting. Results also appeal for an extension of our monitoring system. In particular observing better the vertical structure of temperature and currents would help revising the parameterizations of the mixed layer and baroclinic ocean. Systematizing the use of data to improve model physics is obviously an important step towards progress. The way the data have been used in this paper is not optimal and some choices of parameters may appear arbitrary. The appendix shows that small changes in the method applied to initialize the model with data lead to very different forecasts. Further progress can be obtained if data are assimilated in the coupled model to optimally determine the parameterizations together with the initial conditions (Lee et al, 1999).

Appendix A: Model initialization with sea level data

Section A1 presents the various sea level data sets used to validate the model or to initialize the forecasts. Section A2 describes the procedures of model initialization with sea level data. Section A3 gives some examples of how sensitive the results are to the choice of the sea level data sets. Section A4 gives some examples of the sensitivity to the procedure applied.

A1. Sea level data sets

Data are used either for validating or for initializing the baroclinic anomalies simulated by the ICMs. As the ocean models assume only one vertical mode, the model thermocline depth anomalies are equal to sea level anomalies divided by the model density ratio (0.0057 in CZ). Four data sets providing estimates of sea level or thermocline depth anomalies are used in this study, one comes from TOPEX-Poseidon (TP) altimetric data since October 1992 and three from XBT temperature profiles since 1980.

For each of these data sets, its monthly climatology is computed over as many complete years as possible (4 for TP and 15 for XBT) and removed to determine the anomalies. Data are initially on a 2° by 1° grid on a monthly basis, except for TP along-track data every 10 days which provides a resolution of 0.5° with a maximum inter-track distance of 3°. For model validation, the observed anomalies have all been estimated on the atmospheric model grid (5.625° in longitude and 2° in latitude). For model initialization, except in the Kalman filter experiment described further down, the anomalies have been interpolated in space to match the baroclinic model grid (2° in longitude and 0.5° in latitude) and in time every 10 days to match the model time step.

Hydrographic data provide proxies of sea level or thermocline depth. Three different quantities were derived. The first two, described in Perigaud and Dewitte (1996), are the 20°C isotherm depth (D20) and the oceanic heat content in the upper 400 meters (T400) over 1980-1994 provided by Dr. Smith from Bureau of Meteorology Research

Centre. The third one is the dynamic height at the surface relative to 400m (hdyn) derived from the vertical temperature profiles provided by the BMRC since 1980 with regular updates up to present time (Smith et al, 1995). D20 can be considered as an estimate of the thermocline, T400 and hdyn as proxies of sea level variations. T400, hdyn and TP take into account variations of density over a thicker layer than the upper layer assumed in the baroclinic model (150m) while D20 has a position closer to 150m than 400m. So D20 is directly used for comparison or insertion in the models, whereas the other three estimates are normalized to match the model physics. The normalization factors for the sea level data sets were obtained by matching their level of variability averaged over the Pacific between 15°S and 15°N to D20. First, inter comparison between the three sea level estimates reveals a very good agreement. Over October 1992-March 1998, hdyn and TP have the same level of variability on average over 15°S-15°N (7.2cm) with an rms. difference of 2.9 cm. Over 1980-1994, hdyn and T400 also have the same level of variability (8.1cm) with an rms. difference of 2.1cm. Over this period, T400 is correlated by more than 80% with D20 on average. Its level of variability need to be divided by a coefficient equal to 0.0076 in order to match that of D20. The same normalization factor was applied to all three sea level estimates. Then, the normalized hdyn and D20 thermocline variabilities over 1980-1994 are both 11.5 m on average, with an rms. difference of 5.8 meters. The largest differences are confined in the eastern equatorial Pacific, the Niño3 indices have a variability of 9.4 m for D20 larger than the 8.9 m for hdyn and with an rms. difference of 3.5 meters. NiñoW indices differ by only 2.8 m for a variability of 10.7m for D20 smaller than the 11.1 m for hdyn. In the text, the model thermocline is compared with the normalized hdyn rather than D20 because the latter covers a shorter period, but validation has also been done with D20 over 1980-1994 and lead to the same results about the deficiency of the CZ model in the western Pacific.

When used for model initialization, data accuracy is a much more critical issue than for validation. Indeed, TP anomalies must be referenced to a long time series because the

TP mission covers only a few years since October 1992. So they are corrected with the hdyn anomalous surface averaged over the satellite period. Over October 1992 to January 1995 (the period over which TP data were used to initialize CZ.SL forecasts), this correction is relatively small, ranging between -0.8 cm in the western and +1.3 cm in the eastern equatorial Pacific between 5°S and 5°N (extrema are +3.2 cm at 7°N-140°E and -3.5 cm at 9°S-165°E). The impact of this correction on forecasts is illustrated in Section A3. The data used to initialize the CZ.SL and Tsub.Conv(2.SL) series of this paper are T400, because the hdyn data were not yet available when these series were generated. Section A3 also reports on the sensitivity of forecasts to the choice of the data (T400 or D20) used for initialization.

A2. Initialization procedures of the coupled model with sea level data.

Based on the model assumptions, sea level data contain all the information necessary to initialize the baroclinic fields, i.e. the thermocline depth and the baroclinic currents in the long-wave approximation. The procedure for initializing the CZ.SL forecasts consists in 2 steps. First, sea level data are decomposed into Kelvin and Rossby meridional modes up to the fifth one as explained in Perigaud and Dewitte (1996). Then, the CZ model is integrated in time with the FSU winds and the "observed" baroclinic fields inserted at each time step in order to determine the boundary conditions of the baroclinic model in the western Pacific, as well as the Ekman currents, the SST and the atmospheric fields.

Actually for both steps, alternatives have been tested. For step 1, we tested other methods of assimilating sea level data in the baroclinic mode, including another projection in the interior of the domain, various methods of estimating the baroclinic fields in the vicinity of the western boundary, and Kalman filtering. We verified with the TP time series over October 1992 to January 1995, that the coefficients obtained by the projection method applied in CZ.SL are similar to the ones obtained by Boulanger and Fu (1996) with a different method. This projection recovers well the variability of sea level and zonal currents within 5° of the equator. A second method was developed in order to recover more variability beyond the wave guide. Method 2 consists in projecting sea level data the same way as in method 1, but the Kelvin component is retained and the rest of the sea level is used as the "non-Kelvin" component of the model (see Cane and Patton, 1984). Geostrophic currents derived from the non-Kelvin sea level outside the equatorial band and the Rossby currents within the wave guide are matched by the function proposed by Picaut and Tournier (1991). This second method is the one applied for the series Tsub.Conv (2.SL). Some results obtained by applying either method 1 or method 2 on CZ are reported in Section A4.

Because the model domain covers 124°E to 112°W and has no landmask, the data used in CZ.SL have been extrapolated from the ocean onto the land encountered over the model domain before applying the first step described above. Then the boundary conditions at the westernmost point (124°E) are obtained by verifying the conservation of the meridionally-integrated zonal transport. Because extrapolating data over land is not an accurate solution, other options were tested. One consists in using the "observed" Kelvin and Rossby fields east of 150°E only and running the baroclinic model forced by FSU to determine the values west of 150°E, but this approach needs a matching zone to avoid inconsistencies at the 150°E transition and is not accurate either. Two alternate options that are more reliable have been considered. One is to add a landmask in the model (Boulanger, 1999). The second is to apply an optimal assimilation method, such as the one reported in Appendix A4. In this experiment, along-track TP sea level variations were assimilated with the optimal Kalman filter and smoother developed for the same type of baroclinic model (Fukumori, 1995) to initialize the baroclinic model forced by FSU wind anomalies between October 1992 and December 1993.

An alternative for step 2 consists in using the wind simulated by the model where sea level and FSU wind data have been inserted in a first initialization, to re-force the baroclinic model and re-initialize all the oceanic and atmospheric fields again. This second procedure significantly degrades the baroclinic fields compared to the first one in the CZ case. So rather than gaining in model consistency, the initial conditions retained in the text are the closest to the data, meaning that the CZ.SL series and the results presented below correspond to no more than one initialization of the atmosphere, Ekman and mixed layer with the "observed" baroclinic ocean.

A3. Sensitivity to the choice of sea level data set

Figure A1 compares the SST indices simulated by the CZ model when the TP data referenced to surface to XBT (TOPEX.RefXBT) have been inserted in the model, with the indices obtained when the TP data are variations relative to zero mean over the satellite period (TOPEX.Ref0). The correction affects the Niño3 SST index by less than 0.5°C during the initialization. It does not alter the SST conditions in late 1994 which agree better with the observed index than the CZ.STD. However, it significantly affects the individual forecasts. For example, it makes the forecast initiated in December 1994 colder than observations by 1.3°C in August 1995 and warmer by 0.5°C in July 1996. It is also found that referencing TP to the longer hdyn time series has a strong impact on the forecasts delivered by Tsub.Conv and is a necessary condition to improve the 1997-1998 forecasts delivered by Tsub.Astat (paper in preparation).

Figure A2 compares the results obtained by initializing the CZ model with either the oceanic heat content (T400) or the 20°C isotherm depth (D20). The initial SST series are almost identical (except in 1988), whereas for instance, the model initialized in July 1992 simulates a 12 month-lead SST which is more than 3°C warmer with T400 than with D20. Note that even though the Tsub parameterization of the CZ model is underestimating the impact of a thermocline upwelling on the subsurface temperature, the series initiated with D20 is too cold in 1988. It is also striking that neither D20 nor T400 allow to reproduce the warm event in late 1982-beginning of 1983. This happens even though the difference between D20 and T400 is maximum at that time (D20 is deeper than T400 by as much as 12 m over Niño3 in September 1982). The model thermocline is deeper than D20 by 8 m then. Data errors, as well as inadequacy of the Tsub parameterization, may be responsible for the failure to reproduce the warm event. This issue is discussed below.

A4. Sensitivity to the initialization procedure.

Figure A3 compares the results of the two methods described above when applied to CZ. There is an overall good agreement, except in late 1982 when the model initiated with method 2 simulates a Niño3 SST index which is significantly warmer than with method 1, in better agreement with the observed index. Indeed, data show in late 1982 a strong anomaly in the eastern Pacific beyond 5° of latitude which is not well accounted for by method 1. This result indicates that in addition to model and data errors, the method applied for initializing the model introduces errors as well. Method 1 was developed prior to method 2, when the CZ model was believed to behave like a delayed oscillator and the role of the off-equatorial variability used to be neglected. Method 2 is the most appropriate method for initializing baroclinic models that are based on Cane and Patton (1984). Overall, both methods work reasonably well to reproduce the observed index during the initialization. Figure A3 gives an example initiated in August 1994 where both methods improve the forecasts, but there are cases like in January 1991 (not shown) where the 12 month-lead forecast initiated with method 2 is 2.2°C warmer than observations or than the forecast initiated with method 1 or the CZ.STD forecast. It is also interesting to note that although method2 was applied to Tsub.Conv (2.SL), it does not succeed to initialize the SST to a level in late 1982 as warm as for CZ.SL (compare the dashed lines in Fig 10c and in Fig A3). The weakness of the signal initialized in late 1982 by Tsub.Conv (2.SL) is due to the very slight reduction of the Tsub impact in case of thermocline downwelling (see DP96). Neither this result, nor the fact that D20 is unsuccessful to increase the warming although it is much deeper than T400 (see Fig A2), was expected. It means that the nonlinear relationship between the subsurface temperature, the thermocline depth and the SST over Niño3 is certainly not understood well enough.

Figure A4 compares the results of assimilating TP data with the Kalman filter (TOPEX.KF) or with method 2 as above (TOPEX.Method2). The Niño3 SST indices are significantly different. They differ by about 1°C during the initialization and during the forcasting. All these results certainly highlight the importance of the method chosen to initialize a model with data. The Kalman filter has the big advantage of being optimal. It provides all the components of the baroclinic fields, including the boundary conditions. It also provides error estimates. The assumptions made to estimate the *a priori* model and data errors are the ones described in Fukumori (1995) and are appropriate for the forced context. They are not for the forecasting part however, because the coupled behavior is very different. Optimal methods of assimilation are certainly the preferred choice and the issues relative to error estimations deserve more investigation. Very little is reported in this study which is rather focused on the need to reduce model errors first of all. However this brief overview illustrates the need for further developments and progress in data assimilation methods with ICMs.

In summary, these many experiments support the fact that our failure to improve the CZ forecasts is not primarily due to data errors nor to the inadequacy of the methods applied for initializing the model. At the same time, they illustrate how sensitive the forecasts are to very small changes in the initial conditions and support the need for accurate data and optimal methods of data assimilation. They also suggest that although the initial shock is certainly a source of concern, initialization in a coupled context like in the LDEO2 procedure should be applied to reduce this shock only if it does not degrade the initial conditions beyond the limits of the estimated *a priori* data errors.

Appendix B: Experiments following the LDEO2 procedure

Based on the LDEO2 procedure described in Chen et al (1995a), we performed several experiments in order to try and reproduce their series. Our first experiment consisted in applying the nudging described in their paper between the model wind and the non-detrended FSU starting from the CZ.STD forced conditions of January 1980. Results were different from theirs and we then thought that it was either because the observed winds must be the detrended version of the FSU data set, or because the nudging must be started from rest in January 1964. Then, we used only the detrended winds (as for all the series presented in this paper). Because CZ.CPLIC forecasts are surprisingly sensitive to the starting date of the LDEO2 procedure, experiments with the procedure started either from rest in January 1964 (CPLIC.64), or from the LDEO1 conditions of January 1970 (CPLIC.70), or January 1974 (CPLIC.74) or January 1980 (CPLIC.80) are presented in Fig. B1 and B2. The results are not much sensitive to the starting date during the initialization step (before the vertical line), but they are after. Two predictions arbitrarily chosen many years after the starting date differ by more than 2°C. CPLIC.80 is the CZ.CPLIC series retained in this paper (it is the one that has the best statistics).

CPLIC.80 is not identical to the series presented in Chen et al (1995a; 1997). Neither is the CPLIC.64. One possible explanation is the uncertainty introduced by the interpolation routines applied to match the various grids involved in nudging FSU with the model wind. Because the nudging is applied on the atmospheric grid (Zebiak, personal communication), we estimated FSU data on the model grid prior to applying the LDEO2 procedure by taking the average of the FSU points contained in the atmospheric model grid element weighted by a coefficient function of their distance to the center of the element (interp1). Because the LDEO2 procedure applied to FSU (interp1) winds produces forecasts that are not identical to Chen et al (1995a), a second wind estimate (interp2) was obtained by adding a smoothing on the 4 nearest adjacent points of the grid in latitude and longitude. Results of applying the LDEO2 procedure since January 1980 to (interp1) or (interp2) winds show that the initial series are very similar, but forecasts can be significantly different (Fig. B3 and B4).

The interpolation that delivers the forecasts retained in the paper is (CPLIC.interp1) since 1980. The various CPLIC series developed here are indeed not much different. The Niño3 SST indices all match data similarly well in all cases. As said above, the CZ.CPLIC series is not strictly identical to the one published by Chen et al (1995, 1997a), but it is nevertheless very close (see Figure 3). It was also verified that the CZ.CPLIC wind and thermocline anomalies agree well with the results presented in their papers and that the CZ.STD forecasts are strictly identical to the ones refered to as LDEO1 in the literature.

Finally, we performed the LDEO2 (interp1-detrended-since 1980) experiment where the temporal scheme applied in CZ is replaced by the one applied in Mantua and Battisti (1995). This means that instead of updating the surface current anomalies and upwelling fields after the new SST and the heating fields for the atmosphere have been calculated, the windstress anomaly is used to force the dynamic ocean model and then the surface currents prior to updating the SST and then the atmosphere. We found that the initial indices are not significantly affected by the time integration procedure, but the forecasts are. In summary, these experiments emphasize that even for models as simple as CZ, the coupler which determines the characteristics of the communication between the ocean and the atmosphere in space and time plays a critical role in forecasting.

Acknowledgments

The authors were supported at the Jet Propulsion Laboratory, California Institute of Technology by the National Aeronautics and Space Administration, including Grants 960504 and 959852 with University of California Los Angeles. They thank Drs S. Zebiak and M. Cane from LDEO for providing their code of the coupled ocean-atmosphere model, input files, initialization procedure as well as their time and helpful discussions.

List of references

Barnett T.P., M. Latif, N. Graham, M. Flugel, S. Pazan, and W. White, 1993: ENSO and ENSO related predictability. Part I: Prediction of equatorial Pacific sea surface temperature in a hybrid coupled ocean-atmosphere model. *J. Climate*, **6,** 1545-1566.

Bennett, A.F., B.S. Chua, D.E.Harrison, and M.J.McPhaden, 1998: Generalized inversion of Tropical Atmosphere-Ocean (TAO) Data and a Couple Model of the Tropical Pacific, *J. Climate*, **11**, 1768-1792.

Boulanger JP and L L Fu, 1996: Evidence of boundary reflection of Kelvin wave and first-mode Rossby waves from TOPEX/Poseidon sea level data, *J. Geophys. Res.*, **101**, 16361-16371.

Boulanger J.P., 1999: A study of the 1997-1998 ENSO event. Part I: Model description and validation to data, *Climate Dyn.*, (submitted).

Cassou C. and C. Perigaud, 1998: ENSO simulated with Intermediate Coupled Models and evaluated with observations over 1970-1996. Part II: Role of the off-equatorial ocean discharge-recharge associated with meridional winds, *J. Climate*, (revised).

Chen D., S. E. Zebiak, A. J. Busalacchi and M. A. Cane, 1995a: An improved procedure for El Niño forecasting, *Science*, **269**, 1699-1702.

Chen Y.Q., D.S. Battisti, and E.S.Sarachick, 1995b: A new Ocean Model for studying the tropical Oceanic aspects of ENSO, *J. Phys. Oceanog.*, **25**, 2065-2089.

Chen D., S. E. Zebiak, M. A Cane, and A. J. Busalacchi, 1997: Initialization and Predictability of a coupled ENSO forecast model, *Mon. Wea. Rev.*, **125**, 773-788.

Chen D., M.A. Cane, S.E.Zebiak, and A. Kaplan, 1998: The impact of sea level data assimilation on the Lamont model prediction of the 1997/98 El Niño, *Geophys. Res. Lett.*, **25**, 2837-2840.

Dewitte, B. and C. Perigaud, 1996: El Niño - La Niña events simulated with Cane and Zebiak's model and observed with satellite and in situ data. Part II: Model forced with observations, *J. Climate*, **9**, 1188-1207.

Dewitte B., G. Reverdin and C. Maes, 1998: Vertical structure of an OGCM simulation of the equatorial Pacific in 1985-1994, *J.Phys. Oceanog.,* (in press).

Dewitte B., 1998: "Sensitivity of an intermediate ocean-atmosphere coupled model of the tropical Pacific to its oceanic vertical structure", *J. Climate,* (submitted).

Fischer M., M. Latif, M. Flugel and M. Ji, 1997: The impact of data assimilation on ENSO simulations and predictions, *Mon. Wea. Rev.*, **125**, 819-829.

Fukumori I., 1995: Assimilation of TOPEX sea level measurements with a reduced gravity, shallow-water model of the tropical Pacific Ocean, *J. Geophys. Res.*, **100**, 25,027-25,039.

Gill A.E., 1980: Some simple solutions for heat induced tropical circulation. Q.J.R.Meteor. Soc., 106, 447-462.

Goswami, B.N. and J. Shukla, 1991: Predictability of a coupled ocean-atmosphere model. *J. Climate*, **4**, 1-22.

Ji M., A. Kumar and A. Leetmaa, 1994: An experimental coupled forecast system at the National Meteorological Center. Some early results. *Tellus*, **46A**, 398-418.

Ji M. and A. Leetma, 1997: Impact of data assimilation on Ocean Initialization and El Niño predictions, *Mon. Wea. Rev.*, **125**, 742-753.

Ji M, D.W. Berhinger and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part II: the coupled model. *Mon Wea. Rev.*, **126**, 1022-1034.

Kirtman B. and S.E.Zebiak, 1997: ENSO simulation and prediction with a hybrid coupled model, *Mon Wea. Rev.*, **125**, 2620-2640.

Kleeman R., 1991: A simple model of the atmospheric response to ENSO SST anomalies, *J. Atmos. Sci.,* **48**, 3-18.

Kleeman R., A.M.Moore and N.R.Smith, 1995: Assimilation of sub-surface thermal data into an intermediate tropical coupled ocean-atmosphere model, *Mon.Wea. Rev.*, **123**, 3103-3113.

Latif M., D.Anderson, T.Barnett, M. Cane, R. Kleeman, A. Leetma, J. O'Brien, A.Rosati, E. Schneider, 1998: A review of the predictability and prediction of ENSO, *J. Geophys. Res.*, **103**, 14,375-14,393.

Lee T., J.P. Boulanger, L.L. Fu, A. Foo and R. Giering, 1999: Coupled data assimilation and ENSO prediction: an application to the 97-98 El Niño, *J. Geophys. Res*, (submitted).

Leetma A., and M.Ji, 1989: Operationmal hindcasting of the tropical Pacific, *Dyn. Atmos. Oceans,* **13**, 456-490.

Mantua N.J. and D.S. Battisti, 1995: Aperiodic variability in the Cane -Zebiak coupled ocean-atmosphere model: Ocean atmosphere interactions in the western equatorial Pacific. *J. Climate*, **8**, 2897-2927.

Neelin J.D., and Jin F.F, 1993: Modes of interannual tropical ocean-atmosphere interaction - a unified view. Part II: Analytical results in the weak-coupling limit. *J. Atmos. Sci.* **50:** 3504-3522.

Neelin J.D. and N. Zeng, 1998: The first quasi-equilibrium tropical circulation model formulation, *J. Atmos. Sci.*, (submitted).

Perigaud C. and B. Dewitte, 1996: El Niño - La Niña events simulated with Cane and Zebiak's model and observed with satellite and in situ data. Part I: Model data comparison, *J. Climate*, **9**, 66-84.

Perigaud C., F. Melin and C. Cassou, 1998: ENSO simulated with Intermediate Coupled Models and evaluated with observations over 1970-1996. Part I: Role of the off-equatorial variability, *J. Climate*, (submitted).

Picaut J., and R. Tournier, 1991: Monitoring the 1975-85 equatorial Pacific transport with expendable thermograph, *J. Geophys. Res.*, **96**, 3263-3277.

Rosati A., Miyakoda K. and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a Coupled Model, *Mon. Wea. Rev.*, **125**, 754-772.

Schneider E.K., B.Hunag, Z.Zhu D.G.DeWitt, J.L. Kinter III, B.Kirtman and J.Shukla, 1998: Ocean data assimilation, initialization and predictions of ENSO with a coupled GCM, *Mon. Wea. Rev.* (in press).

Smith N.R., J.E. Blomley and G. Meyers, 1995: An improved system for tropical ocean sub-surface temperature analyses, *J. Atmos. Oceanic. Technol.*, 219-256.

Syu H.H. and J.D. Neelin, 1998: ENSO in a hybrid coupled model: sensitivity to physical parameterizations and prediction with piggyback data assimilation, *Clim. Dyn.*, (submitted).

Torrence C. and G.P.Compo, 1998:A Practical Guide to Wavelet Analysis, *Bull. Amer. Meteor. Soc.*, **79**, 61-78.

Wang B. and and T. Li, 1993: A simple tropical atmosphere model of relevance to short-term climate variations. *J. Atmos. Sci.*, 260-284.

Wunsch C., 1999: The interpretation of Short Climate Records, with Comments on the North Atlantic and Southern Oscillations, *Bull. Am. Meteor. Soc.*, **80**, 245-256.

Xue Y., M.A. Cane and S.E.Zebiak, 1997: Predictability of a coupled Model of ENSO using Singular Vector Analysis. Part II:Optimal Growth and forecast Skill, *Mon. Wea. Rev.*, **125**, 2074-2093.

Zebiak S.E., 1986: Atmospheric Convergence Feedback in a simple model for El Niño, *Mon Wea. Rev.*, **115**, 2262-2278.

Figure Caption

Figure 1: Statistics derived from CZ or sine forecasts and observations. **1ab)** Correlation and r.m.s difference in (°C) between the observed and the forecasted Niño3 SST indices as a function of lead-time in months. The plain line corresponds to CZ forecasts and the dot-dashed one to the persistence. **1cd)** are the histograms of the period determined by best fitting sine functions to the series of two-year long segments of observed (left) or CZ forecasted (right) indices. **1ef)** correspond to 1ab) for sine fits performed with a period equal to 2 (dash) or 4 (plain) years, the fit being done during the two years of interest (i.e. this fit is not a forecast). **1ghij)** correspond to 1ab) where the forecasts are either CZ (dotted) or 4 year sine functions (plain) fitted to the data during the 4 years prior to the forecasts. **1kl)** correspond to 3 year sine forecasts that initially fit the observed index and have a fixed amplitude of 2°C and an initial warming trend (plain), or have a random amplitude between 0 and 2°C and an initial trend which has a positive (dashed), negative (dotted) or random (dot-dashed) sign. Statistics for 1abcdef) are computed over 1972-1995, those for 1gh) over 1974-1985, for 1ij) over 1984-1995, for 1k) over 1980-1995 and for 1l) over 1970-1997. The 95% levels of confidence are indicated by the bars in the right top corner of each correlation panel and the levels of observed variability by the plain thin lines in each error panel.

Figure 2: Time series of Niño3 SST index between January 1981 and 1993 from data (plain), or from CZ during 12 month-long forecasts (dotted). Top panel corrresponds to the first year of forecasts (with the initial conditions in dashed). Bottom panel corresponds to the second year of forecasts (with the series for lead-time=12 months in dashed). Fig2ab corresponds to the CZ.STD series, Fig 2cd to CZ.SL.

Figure 3: Same as Figure 2 for SST Niño3 (Fig.3ab) and H NiñoW (Fig.3 cd) indices simulated by CZ.CPLIC.

Figure 4: Maps of RMS variability of the thermocline depth in meter over the period 1980-1993 **a)** derived from data or initialized by **b)** the LDEO1 procedure (CZ.STD), **c)**

the data assimilation method1 described in the Appendix (CZ.SL), **d)** the LDEO2 procedure (CZ.CPLIC).

Figure 5: Thermocline depth in meters as a function of longitude on average between 1°S and 1°N derived from observations (dashed), CZ.SL (dotted) or CZ.CPLIC (plain). Plots in abcd) correspond to lead-times equal to 1 month, plots in ef) to 6 month. Plots have been averaged in time over the period indicated in the title of each panel.

Figure 6: Zonal wind stress anomalies observed by FSU data (dashed) or by Astat wind (dotted) or simulated by the LDEO2 procedure for initializing the forecasts (plain) as a function of longitude along the equator. Model and data are averaged in time over the months indicated in the title of each panel.

Figure 7: Comparison between observations (OBS) and initial conditions obtained by the LDEO1 (CZ.STD) or LDEO2 procedures (CZ.CPLIC) as a function of longitude for off-equatorial bands. **ab)** Zonal wind stress averaged between 8°N and 12°N and during 5 months centered on January 1983 or 1992. The values for the CZ.STD plain thin curve have been multiplied by 0.45 which corresponds to the value of the nudging coefficient applied in LDEO2 at these latitudes. **cd)** Thermocline depth averaged between 6°N and 10°N and between July and October 1983 or 1992. The dashed line corresponds to the CZ model with a friction equal to 6 months instead of 30 months in CZ.STD or CZ.CPLIC.

Figure 8: Time series of the Niño3 SST and the NiñoW thermocline between January 1981 and 1993 from data (plain), or from 6 month lead-time forecasts simulated by CZ.STD (dashed) or by Tsub.Conv (dotted).

Figure 9: Variability of the zonal wind stress anomalies over 1981-1994 in Dyn/cm^2. Top map is for observations (Astat wind), the bottom three maps are for forecasts with a 6 month lead-time, recpetively for CZ, CZ.CPLIC and Tsub.Conv.

Figure 10: Time series of Niño3 SST index between January 1981 and 1993 from data (plain), or from the Tsub.Conv model during 12-month of forecasts (dotted) during the

first year. The dashed-line represents the initial conditions before forecasting. See text for nomenclature.

Figure 11: Same as Figure 10 for Tsub.Astat. Note that the two time series in **b)** and **c)** cover a longer period than **a)**, up to December 1997.

Figure 12: Maps of SST anomalies in Degree Celsius derived from data (top), Tsub.Astat (middle) or CZ.CPLIC (bottom) forecasts with a 12 month lead-time. Left maps are averaged anomalies between March and September 1987, middle between July and December 1988 and right between January and July 1992.

Figure 13: Same as Figure 12 for the zonal wind stress anomaly in Dyn/cm^2.

Figure 14: Same as Figure 12 for the meridional wind stress anomaly in Dyn/cm^2.

Figure 15: Thermocline anomaly in meters zonally averaged over the Pacific as a function of latitude and years between January 1981 and December 1997. Left diagram corresponds to observations, middle to Tsub.Astat and right to CZ.CPLIC ensemble forecasts for lead-times between month 1 and 12.

Figure A:

Nino3 SST indices as a function of time. In all panels, the observed index is represented by the thick plain line, the CZ.STD index is represented by the thin plain line, and the vertical bar separates the period of initialization from the period of forecasting. The dotted and dashed lines correspond to the series simulated by the CZ model initialized with sea level data, following the tests and nomenclatures described in Appendix A.

Figure B:

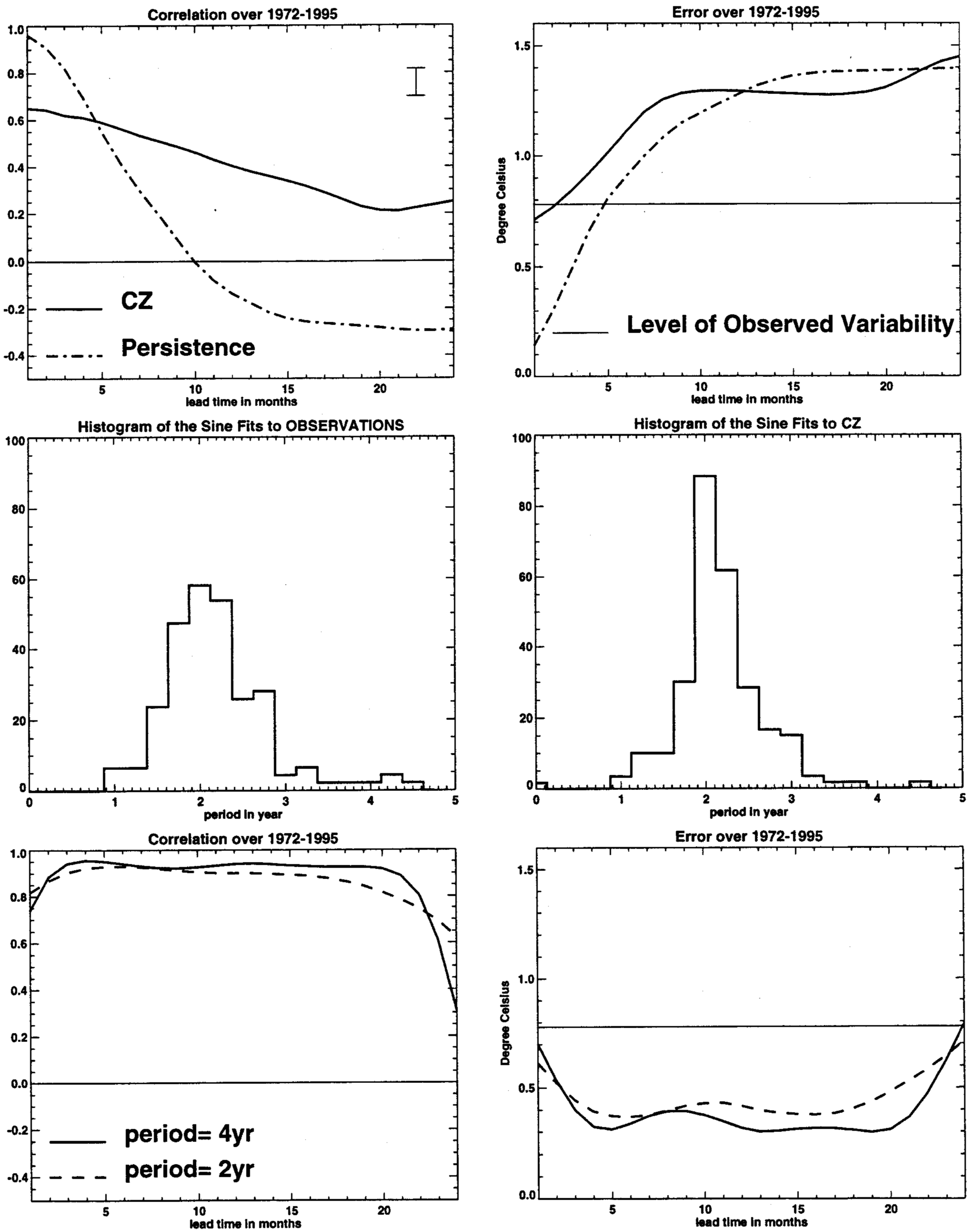
Nino3 SST indices as a function of time. In all panels, the observed index is represented by the thick plain line and the vertical bar separates the period of initialization from the period of forecasting. The curves correspond to the series of indices simulated by the CZ model initialized with the LDEO2 procedure for the various tests described in Appendix B.

Table1: Identification of the Tsub.Conv experiments

Name	Time decay month	Convect weight %	Initialization procedure	Wind forcing for generating ICs
Tsub.Conv (1)	3	100	LDEO1	FSU
Tsub.Conv (2)	24	10	LDEO1	FSU
Tsub.Con (2.SL)	24	10	SL (method 2)	FSU
Tsub.Conv (3)	6	100	LDEO1	Astat (first SVD)
Tsub.Conv (3.CPLIC)	6	100	LDEO2	Astat (first SVD)

Table2: Identification of the Tsub.Astat experiments

Name	Time decay month	Coupl. Coeff.	Initialization procedure	Wind forcing for generating ICs
Tsub.Astat (1)	6	1.0	LDEO1	Astat (first SVD)
Tsub.Astat (2)	6	1.0	LDEO1	Astat (7 first SVD)
Tsub.Astat (3)	6	1.2	LDEO1	Astat (7 first SVD)



Correlation over 1972-1995
CZ
Persistence
lead time in months
Error over 1972-1995
Degree Celsius
Level of Observed Variability
lead time in months
Histogram of the Sine Fits to OBSERVATIONS
period in year
Histogram of the Sine Fits to CZ
period in year
Correlation over 1972-1995
period= 4yr
period= 2yr
lead time in months
Error over 1972-1995
Degree Celsius
lead time in months

Fig.1abcdef

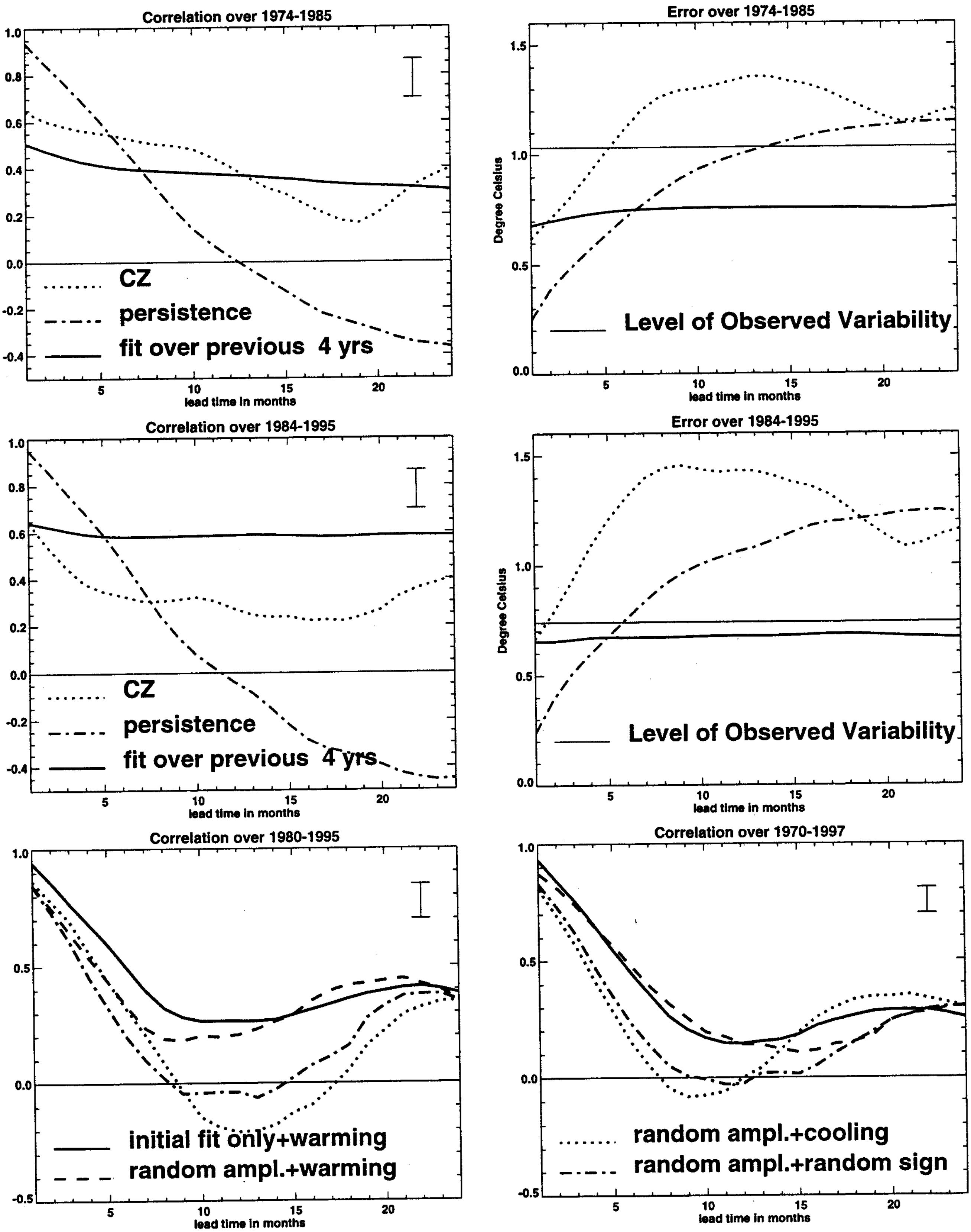
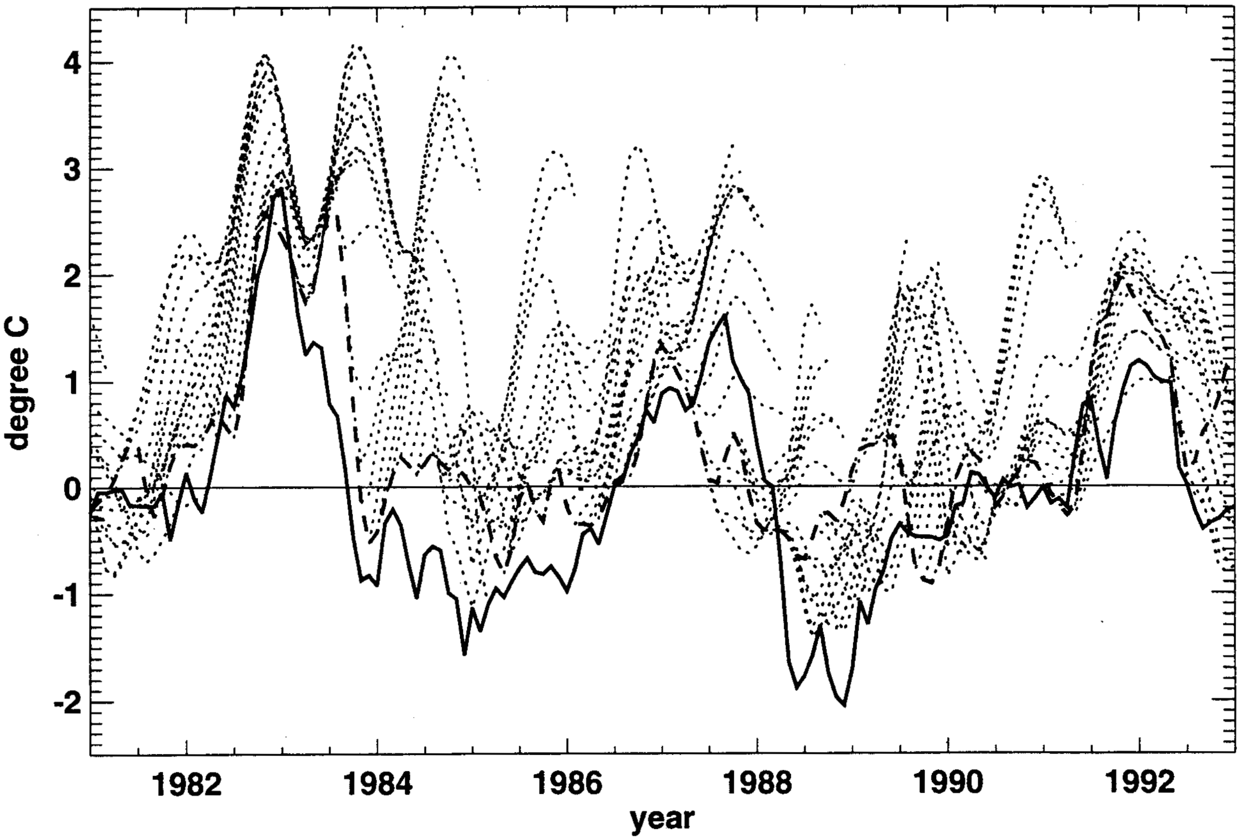

Correlation over 1974-1985
CZ
persistence
fit over previous 4 yrs
lead time in months
Error over 1974-1985
Degree Celsius
Level of Observed Variability
lead time in months
Correlation over 1984-1995
CZ
persistence
fit over previous 4 yrs
lead time in months
Error over 1984-1995
Degree Celsius
Level of Observed Variability
lead time in months
Correlation over 1980-1995
initial fit only+warming
random ampl.+warming
lead time in months
Correlation over 1970-1997
random ampl.+cooling
random ampl.+random sign
lead time in months

Fig.1ghijkl

SST CZ.STD: first 12 months of forecast



SST CZ.STD: months 13 to 24 of forecast

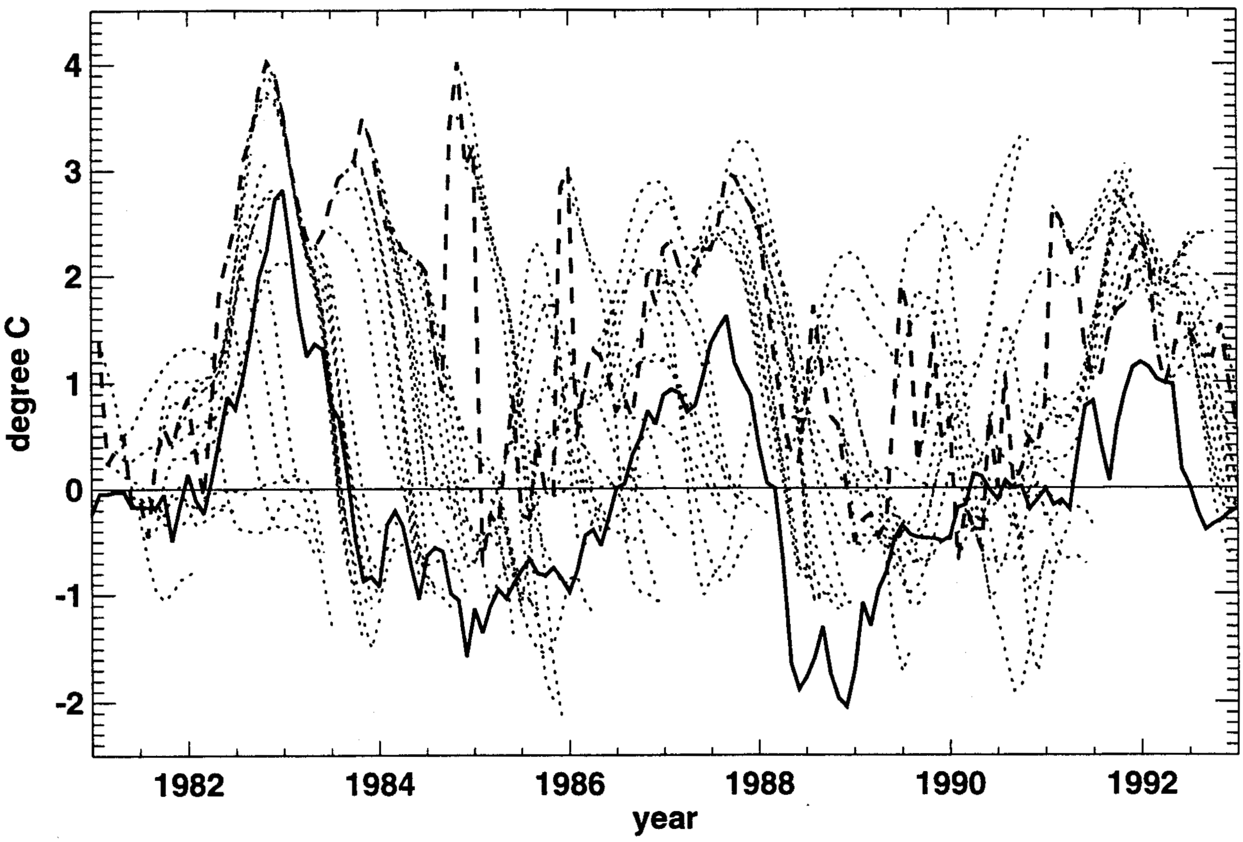
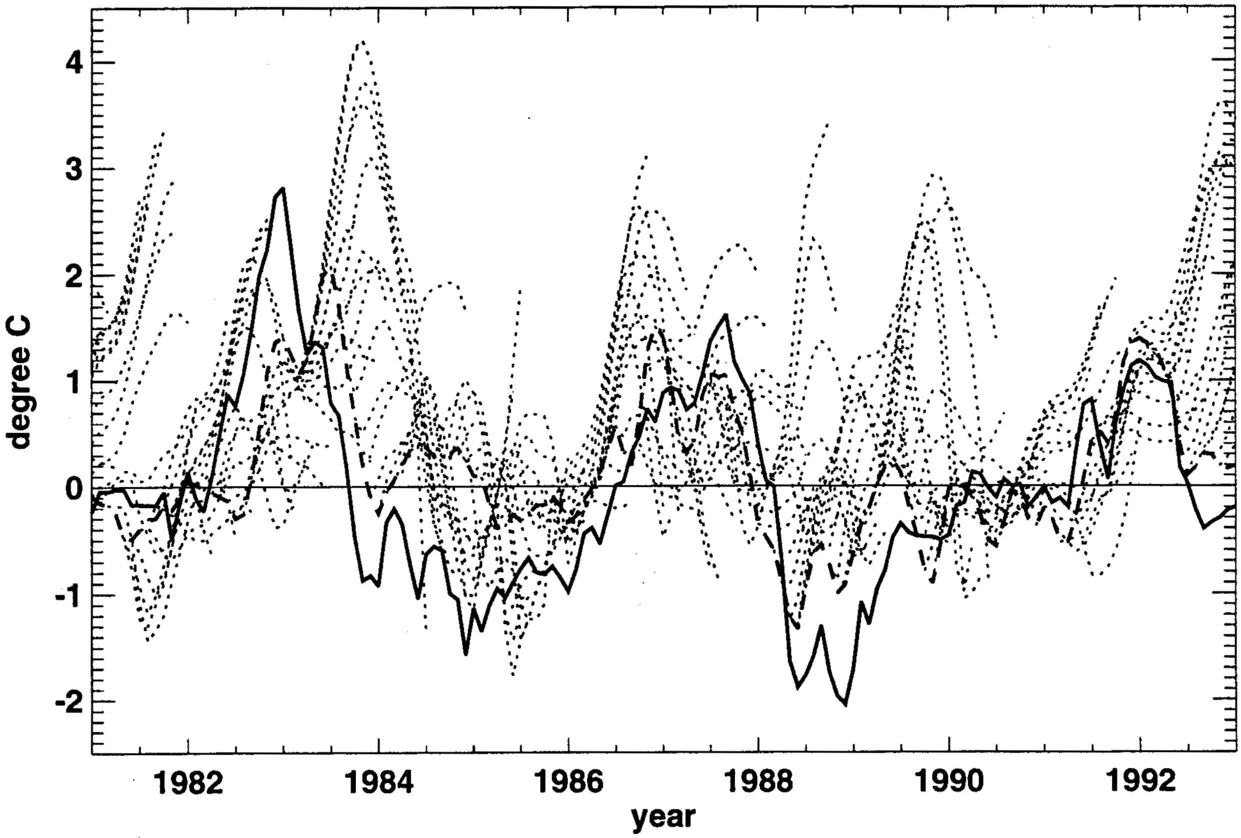


Fig.2ab

SST CZ.SL: first 12 months of forecast



SST CZ.SL: months 13 to 24 of forecast

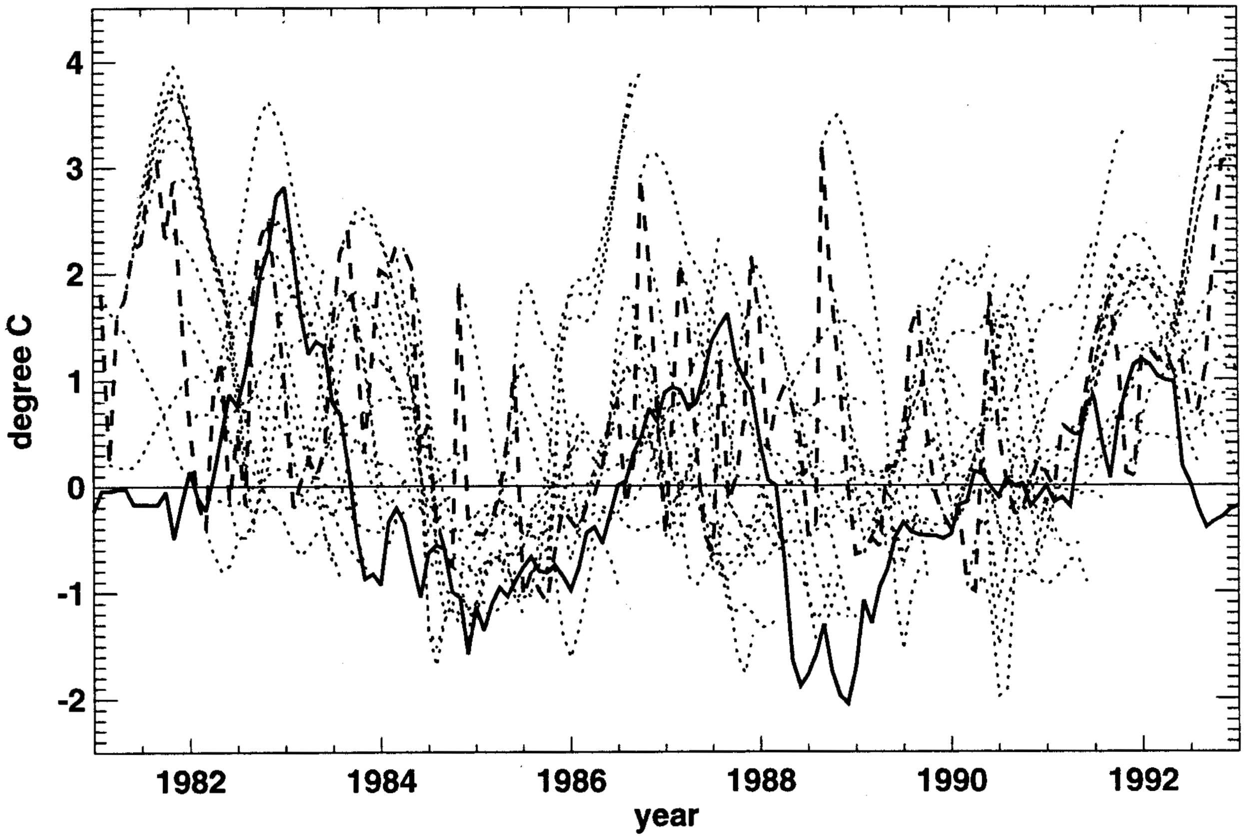
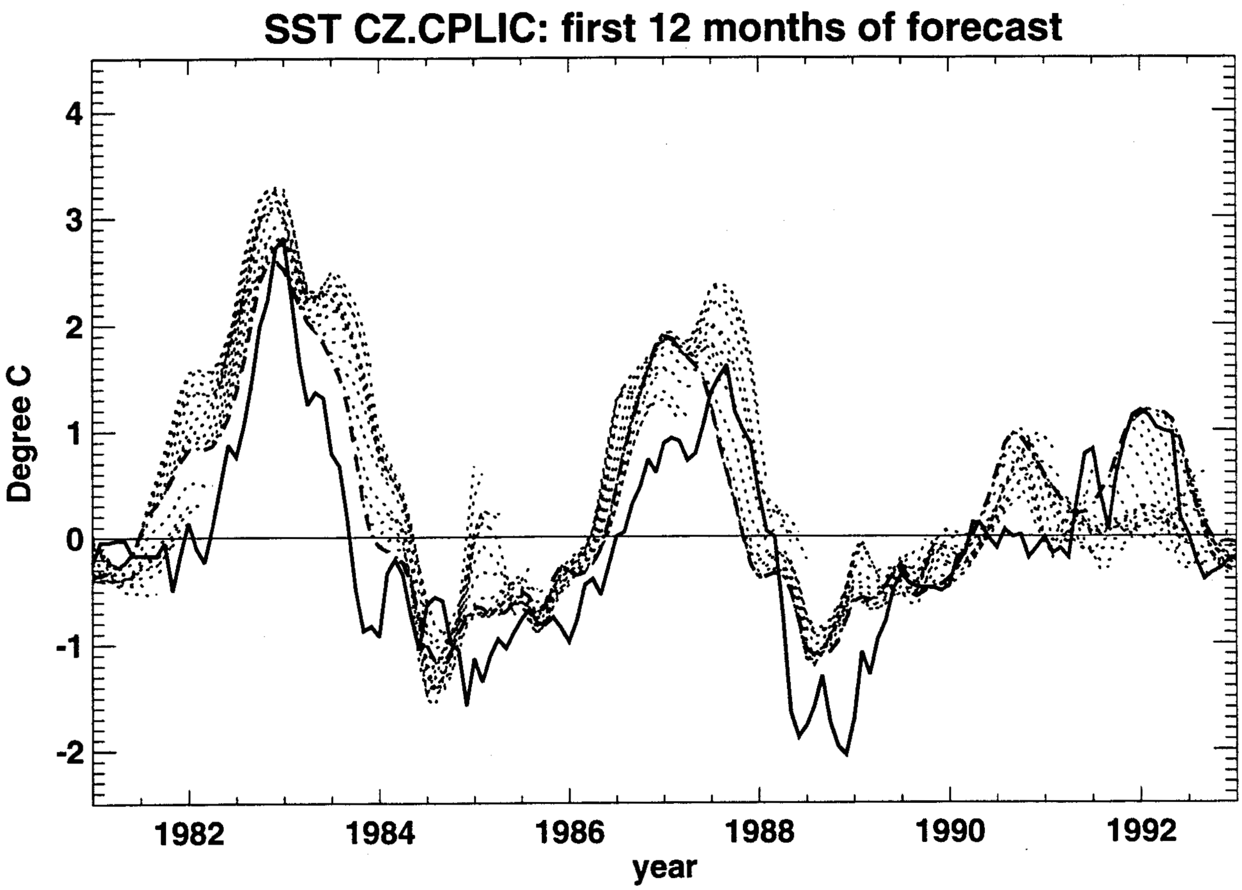


Fig.2cd


SST CZ.CPLIC: first 12 months of forecast
Degree C
4
3
2
1
0
-1
-2
1982
1984
1986
1988
1990
1992
year

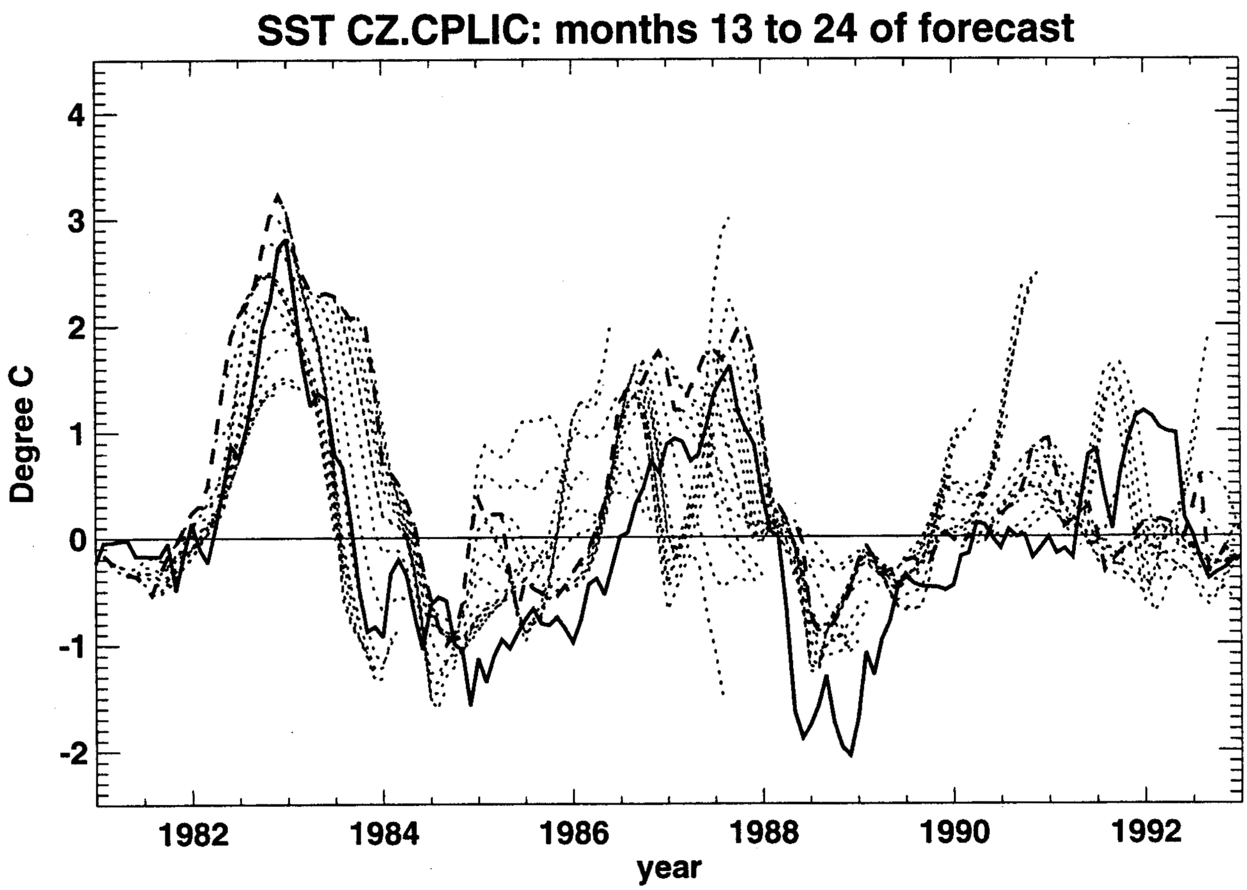
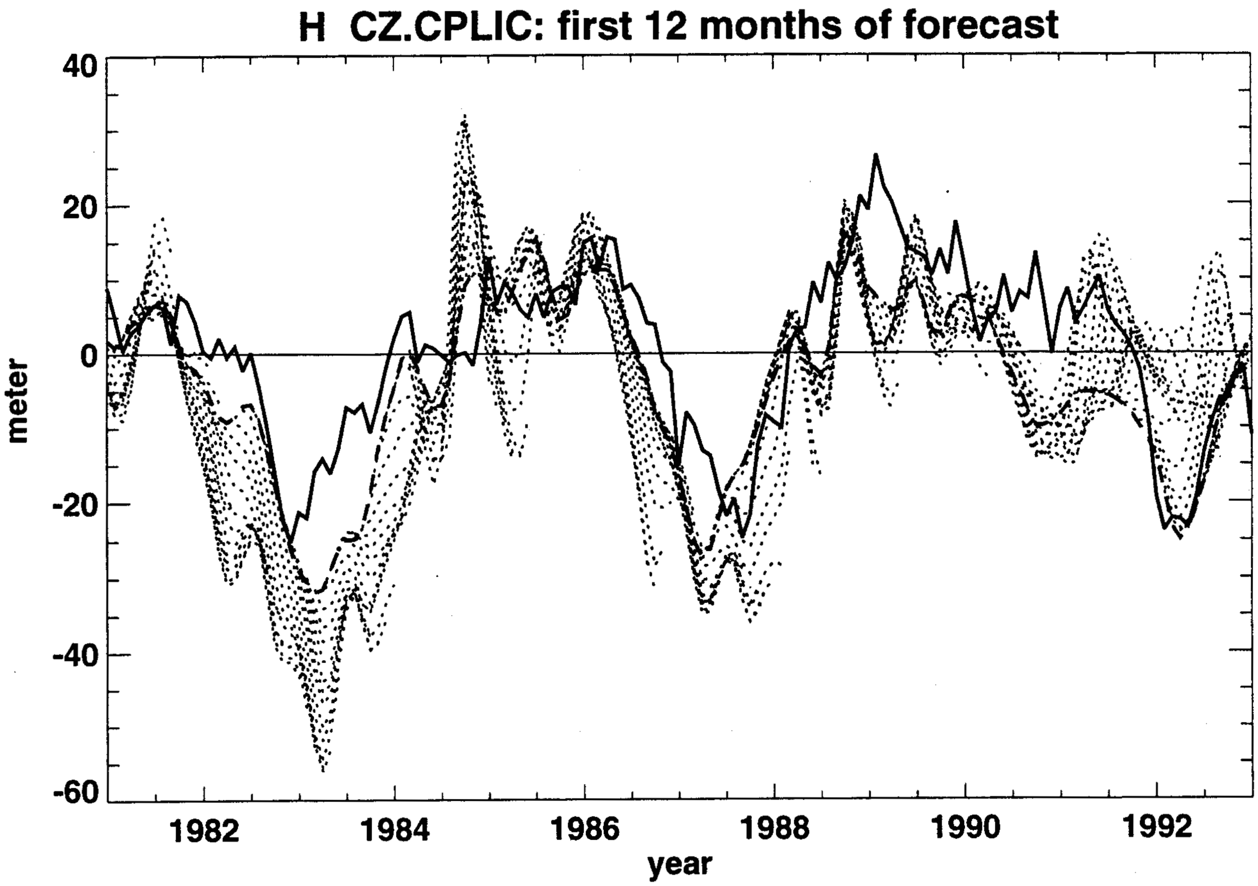

SST CZ.CPLIC: months 13 to 24 of forecast
Degree C
4
3
2
1
0
-1
-2
1982
1984
1986
1988
1990
1992
year

Fig.3ab


H CZ.CPLIC: first 12 months of forecast
40
20
0
-20
-40
-60
meter
1982
1984
1986
1988
1990
1992
year

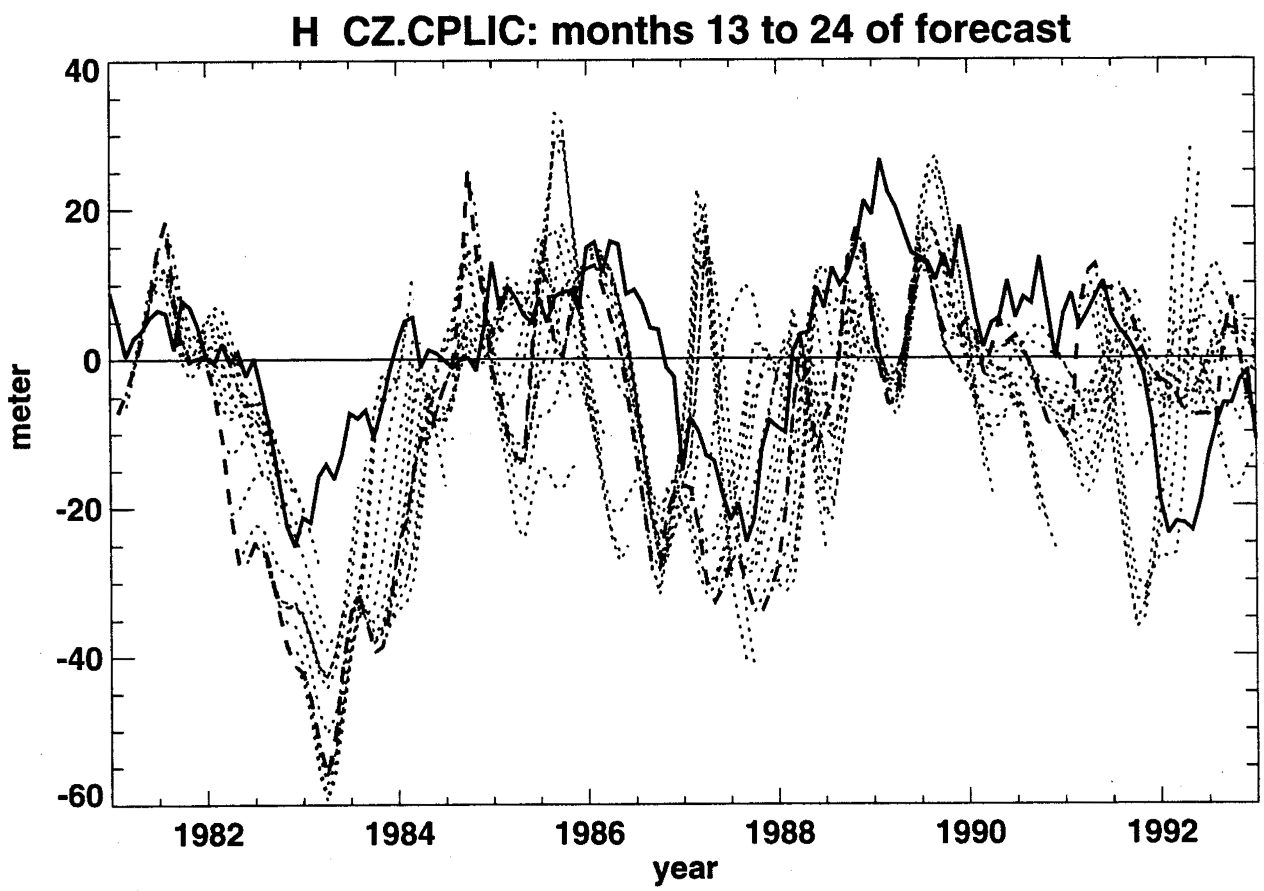

H CZ.CPLIC: months 13 to 24 of forecast
40
20
0
-20
-40
-60
meter
1982
1984
1986
1988
1990
1992
year

Fig.3cd

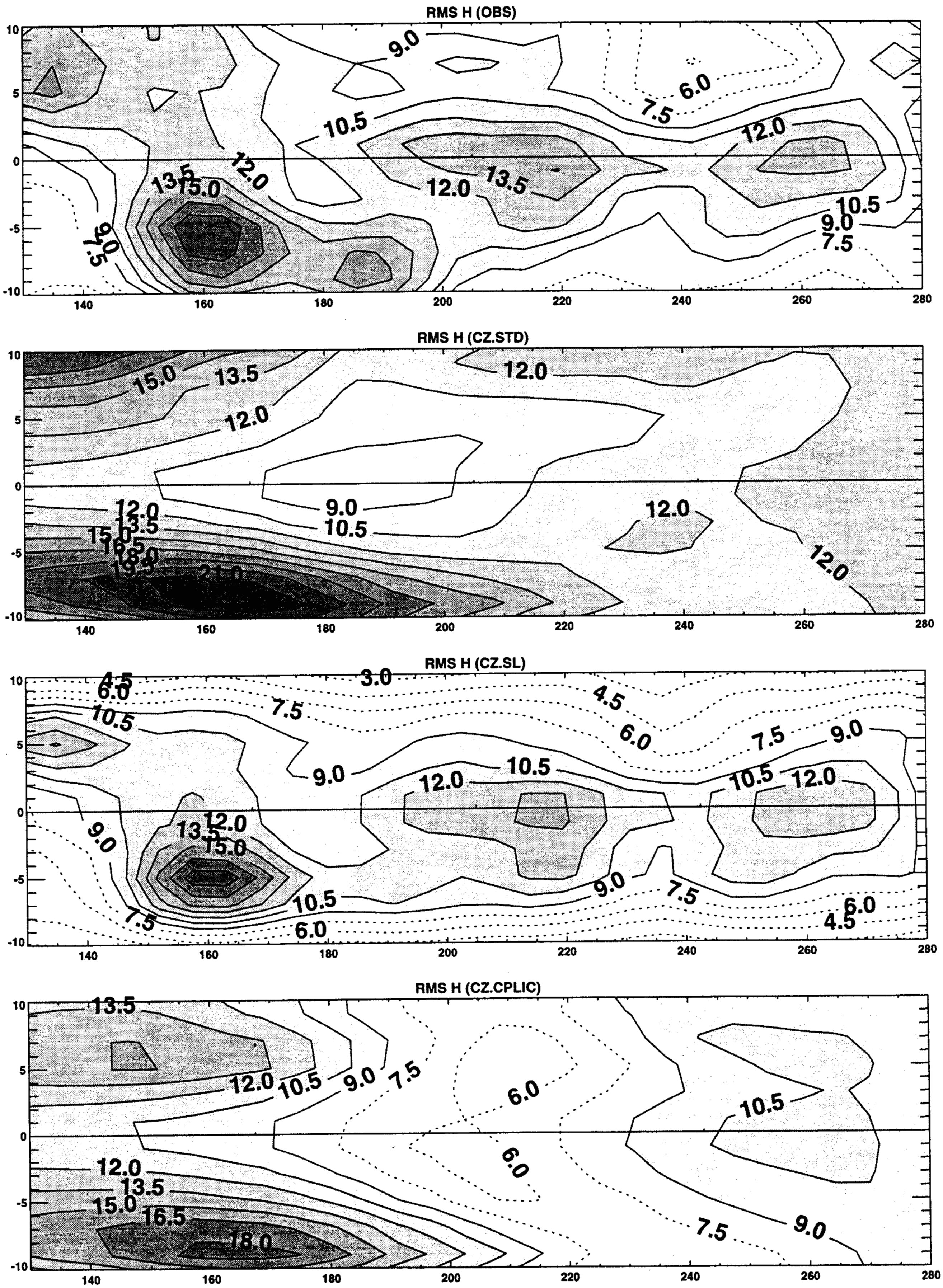

RMS H (OBS)
RMS H (CZ.STD)
RMS H (CZ.SL)
RMS H (CZ.CPLIC)
10
5
0
-5
-10
140
160
180
200
220
240
260
280

Fig.4abcd

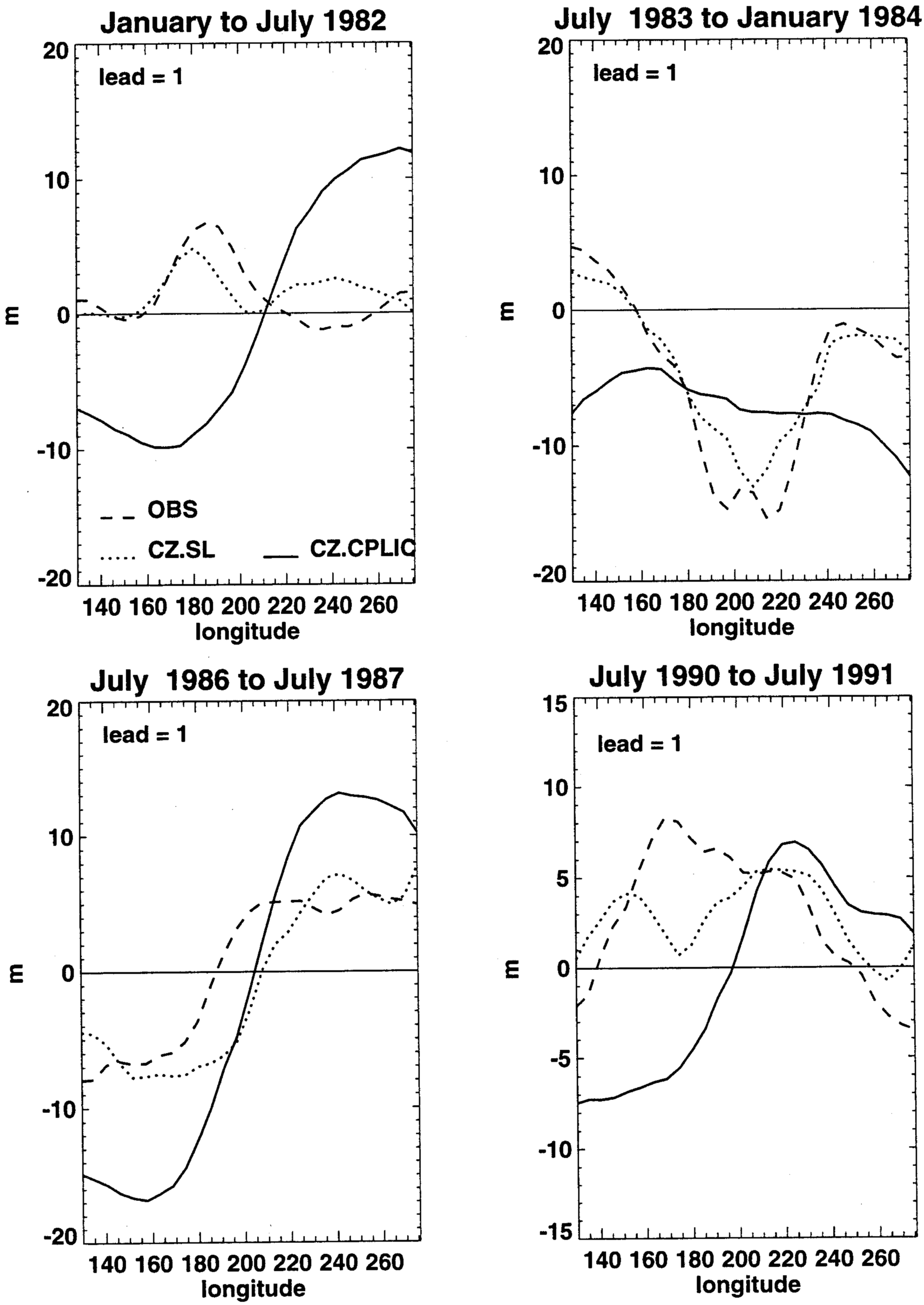

January to July 1982
July 1983 to January 1984
July 1986 to July 1987
July 1990 to July 1991
lead = 1
OBS
CZ.SL
CZ.CPLIC
m
longitude
20
10
0
-10
-20
15
5
-5
-15
140 160 180 200 220 240 260

Fig.5abcc

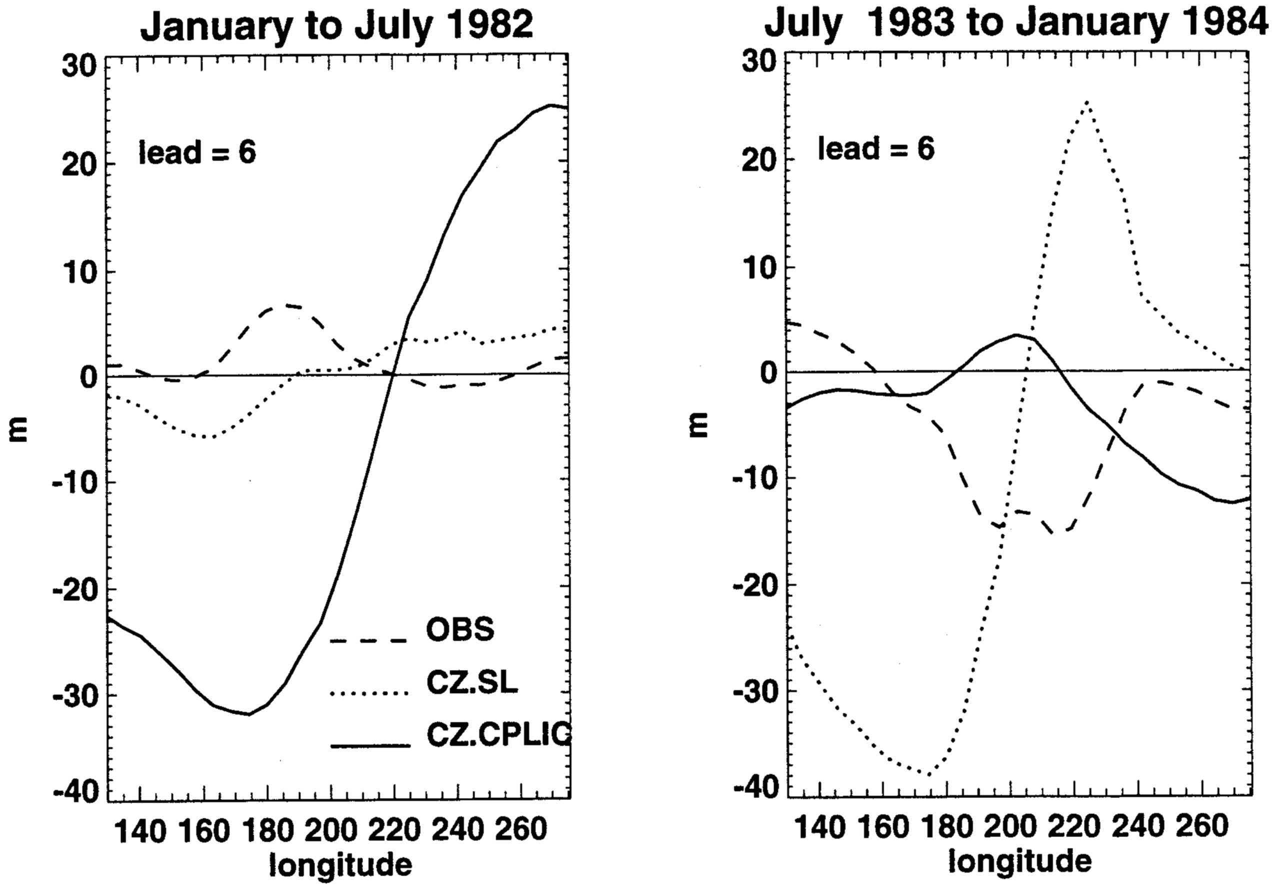
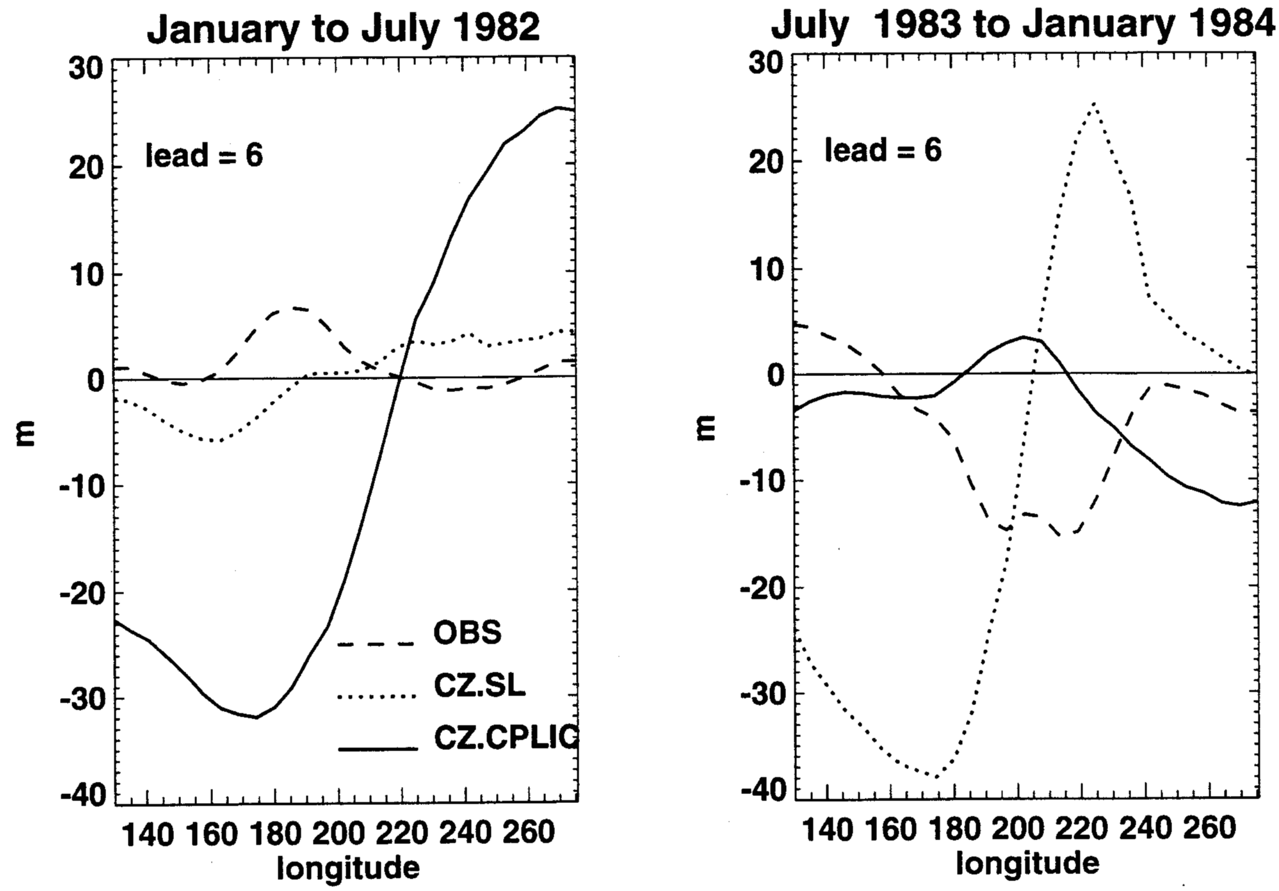

January to July 1982
lead = 6
OBS
CZ.SL
CZ.CPLIC
m
longitude
30
20
10
0
-10
-20
-30
-40
140 160 180 200 220 240 260
July 1983 to January 1984
lead = 6
m
longitude
30
20
10
0
-10
-20
-30
-40
140 160 180 200 220 240 260

Fig.5ef



January to July 1982
lead = 6
OBS
CZ.SL
CZ.CPLIC
m
longitude
30
20
10
0
-10
-20
-30
-40
140 160 180 200 220 240 260
July 1983 to January 1984
lead = 6
m
longitude
30
20
10
0
-10
-20
-30
-40
140 160 180 200 220 240 260

Fig.5ef

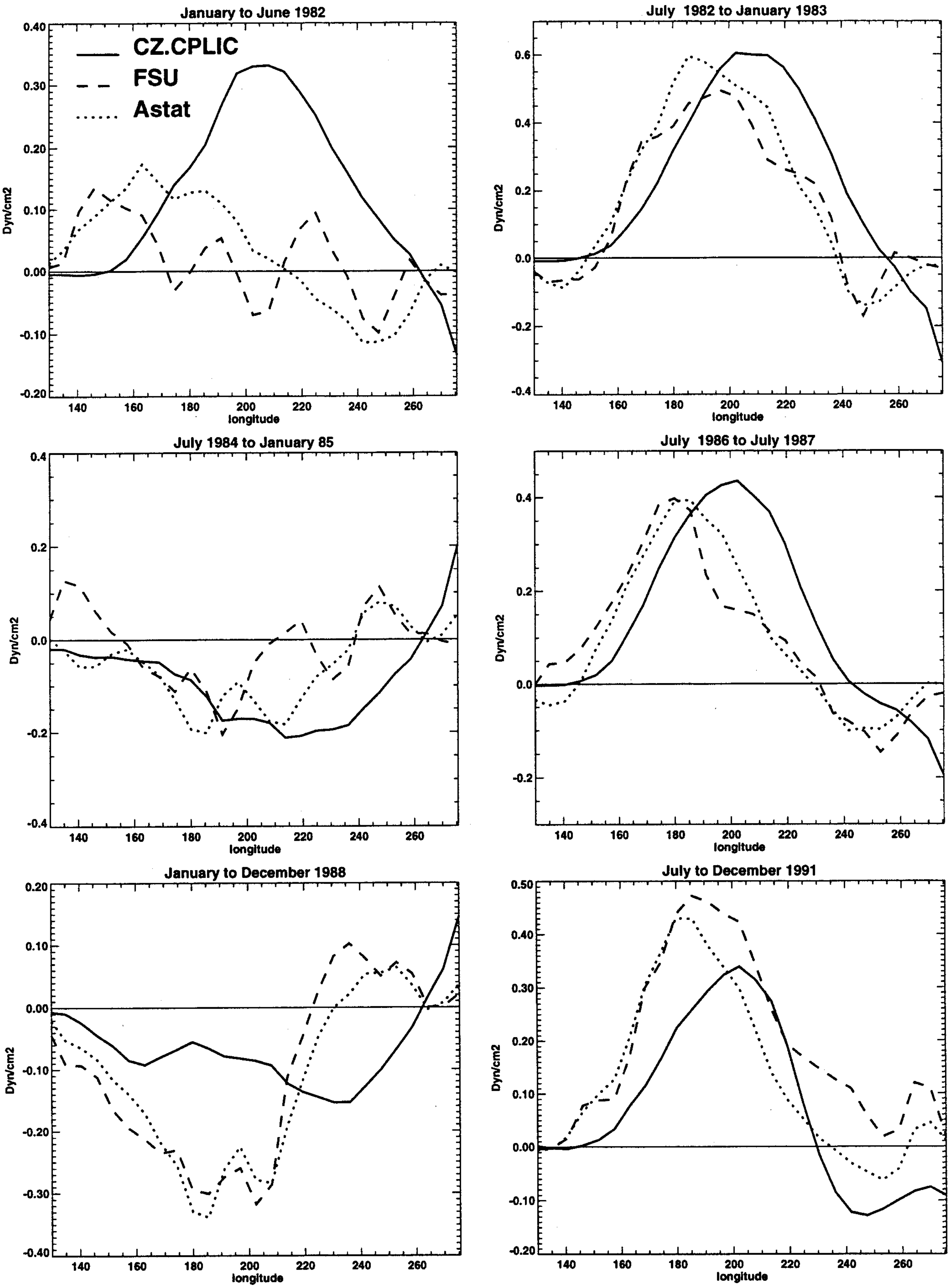
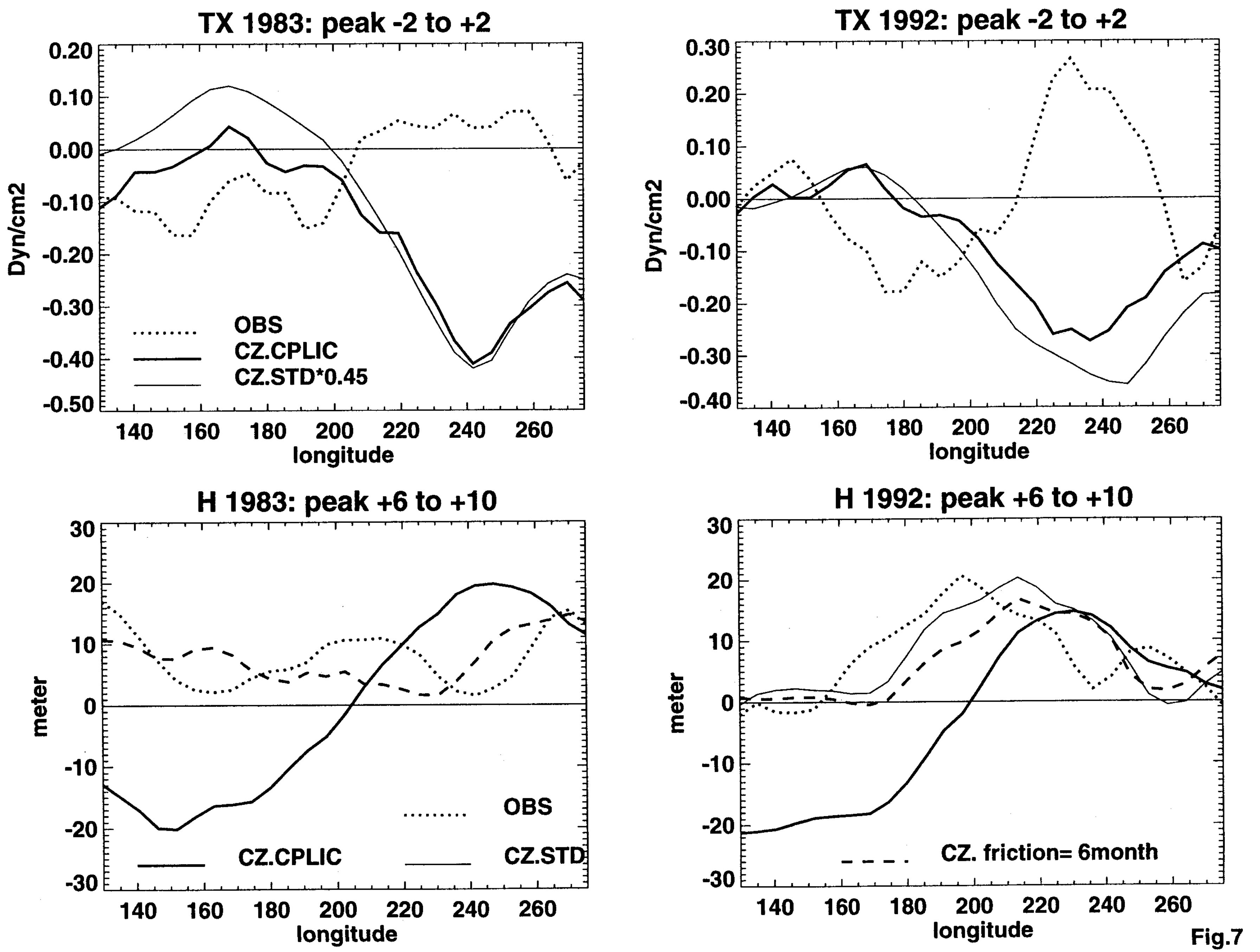

January to June 1982
CZ.CPLIC
FSU
Astat
Dyn/cm2
longitude
July 1982 to January 1983
July 1984 to January 85
July 1986 to July 1987
January to December 1988
July to December 1991
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
0.6
0.4
0.2
0.0
-0.2
-0.4
0.50
140
160
180
200
220
240
260

Fig.6



TX 1983: peak -2 to +2
Dyn/cm2
longitude
OBS
CZ.CPLIC
CZ.STD*0.45
TX 1992: peak -2 to +2
H 1983: peak +6 to +10
meter
CZ.STD
H 1992: peak +6 to +10
CZ. friction= 6month

Fig.7

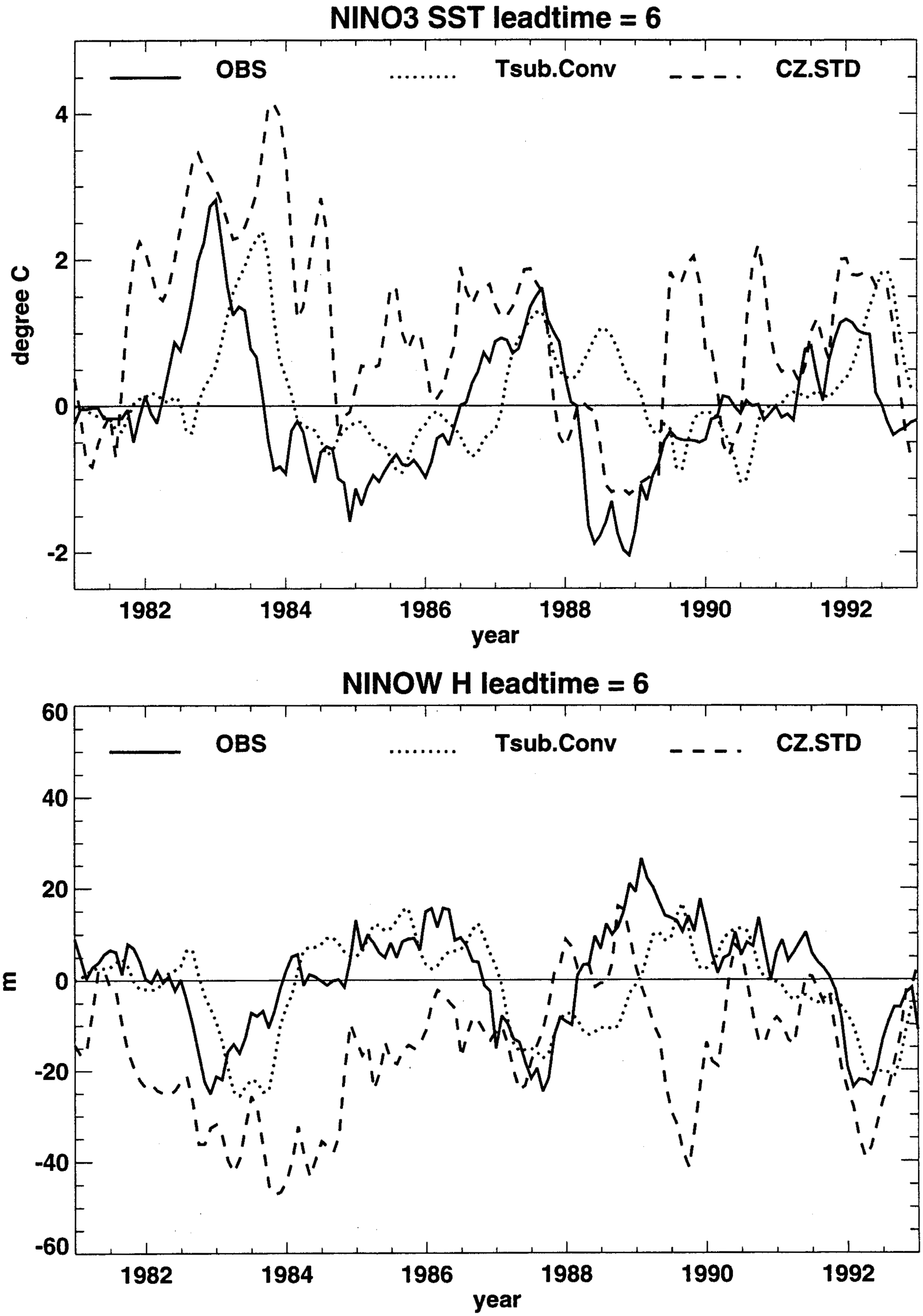

NINO3 SST leadtime = 6
OBS
Tsub.Conv
CZ.STD
degree C
4
2
0
-2
1982
1984
1986
1988
1990
1992
year
NINOW H leadtime = 6
OBS
Tsub.Conv
CZ.STD
m
60
40
20
0
-20
-40
-60
1982
1984
1986
1988
1990
1992
year

Fig.8ab

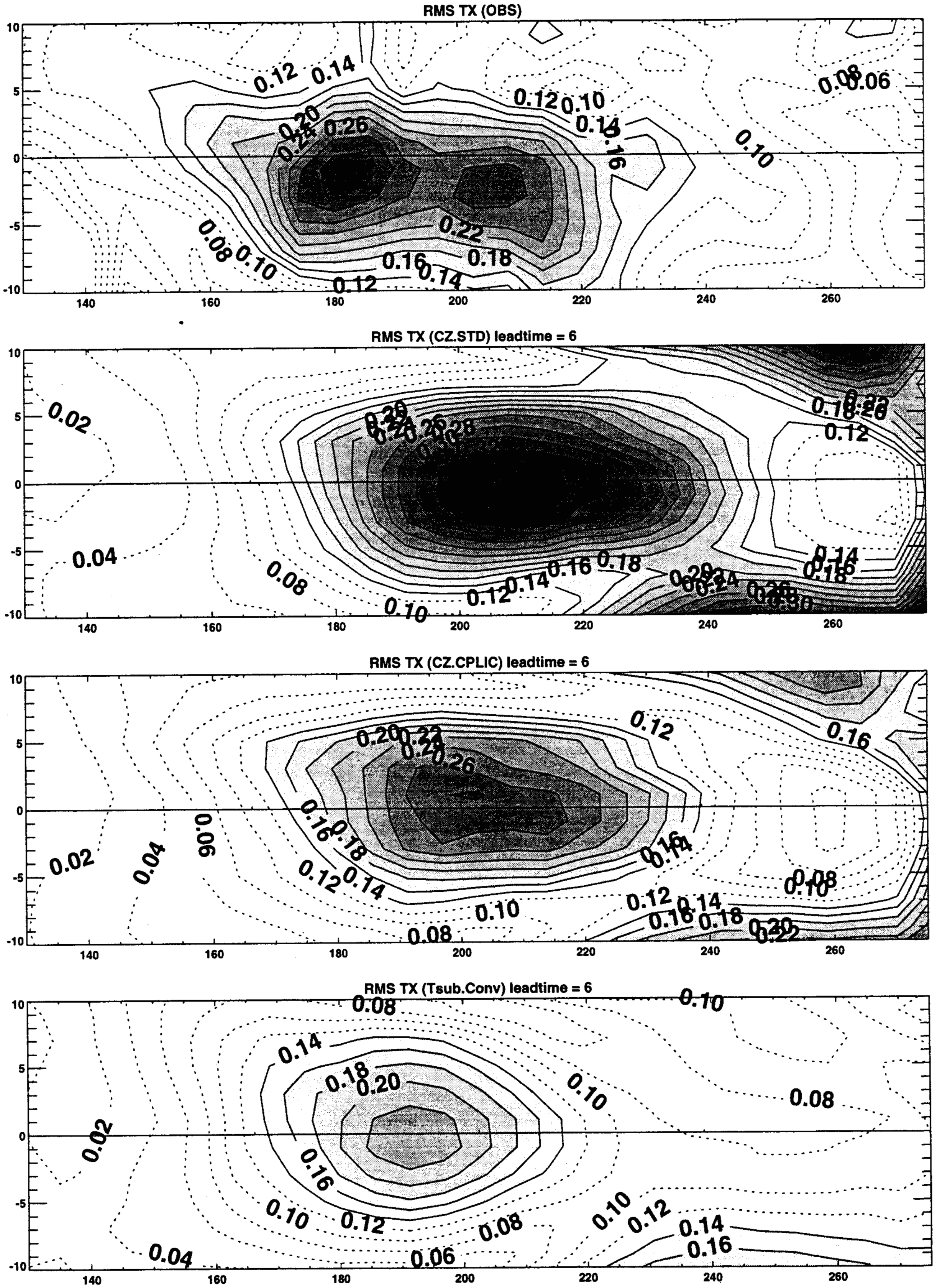
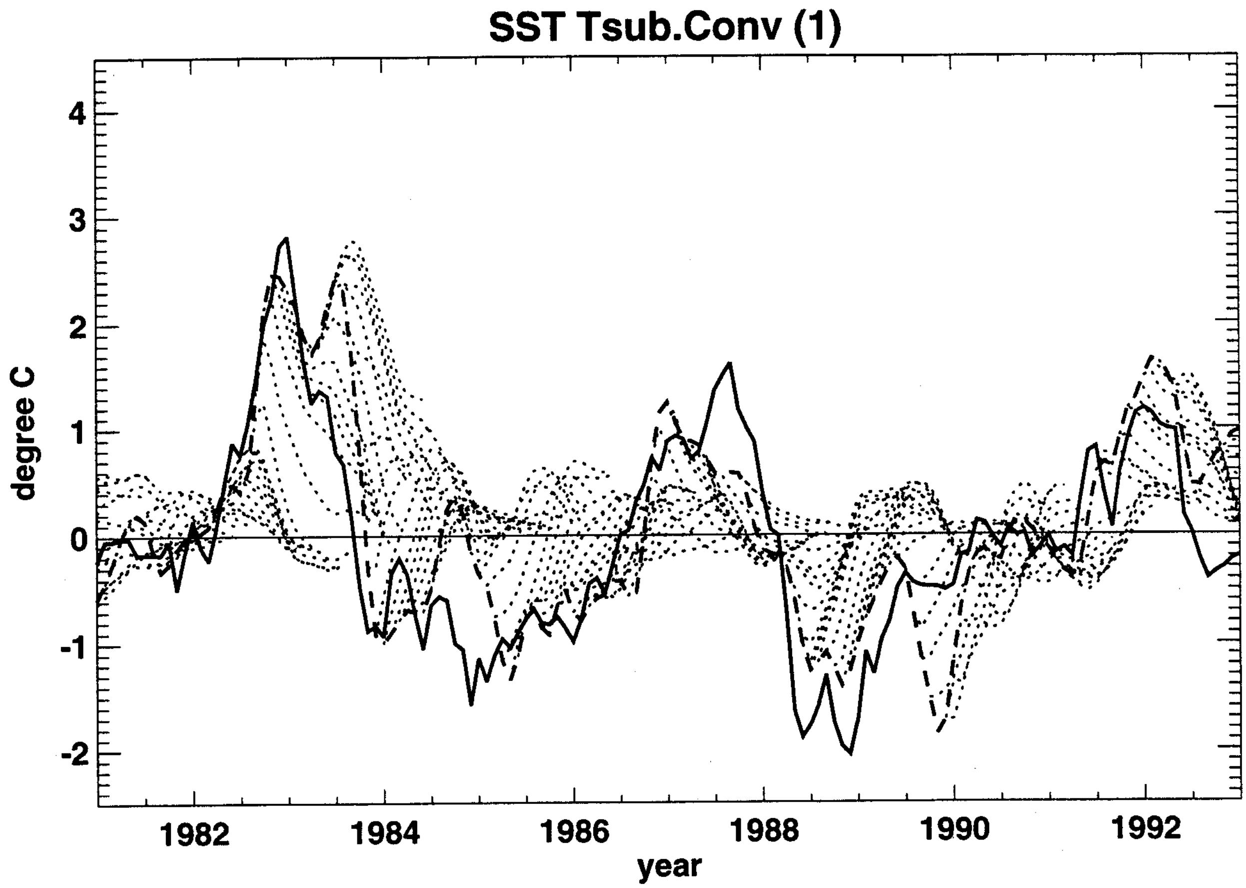

RMS TX (OBS)
RMS TX (CZ.STD) leadtime = 6
RMS TX (CZ.CPLIC) leadtime = 6
RMS TX (Tsub.Conv) leadtime = 6

Fig.9


SST Tsub.Conv (1)
degree C
4
3
2
1
0
-1
-2
1982
1984
1986
1988
1990
1992
year

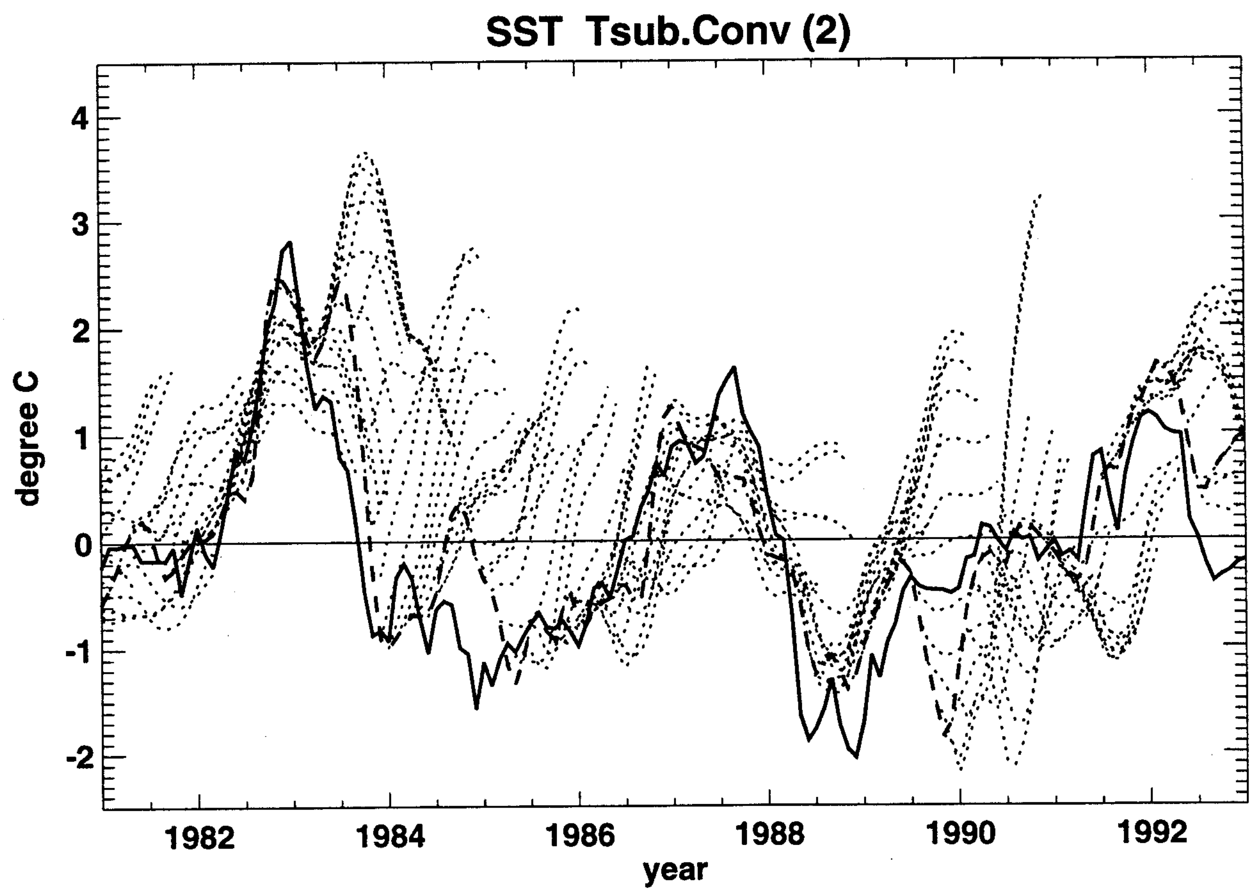

SST Tsub.Conv (2)
degree C
4
3
2
1
0
-1
-2
1982
1984
1986
1988
1990
1992
year

Fig.10ab

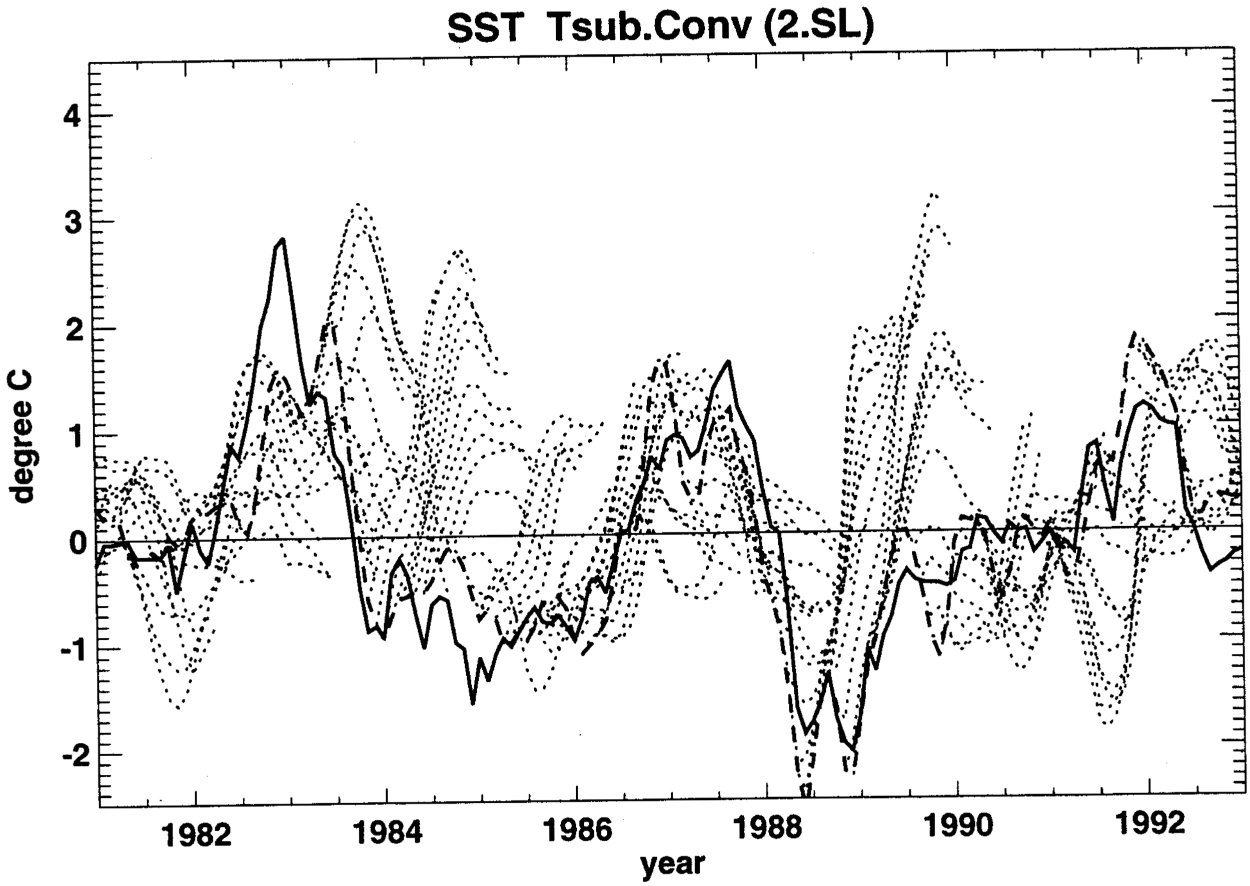
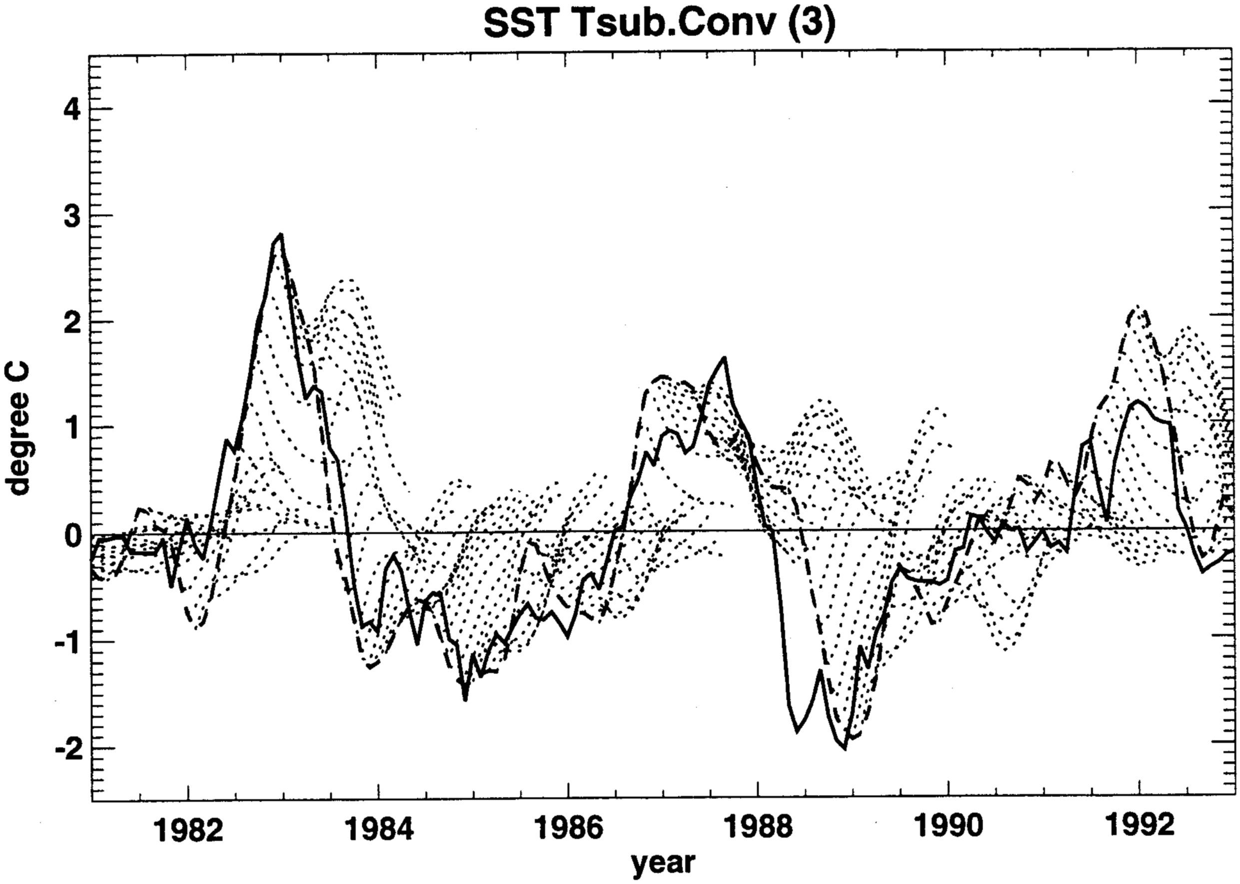

SST Tsub.Conv (2.SL)
degree C
4
3
2
1
0
-1
-2
1982
1984
1986
1988
1990
1992
year

Fig.10c


SST Tsub.Conv (3)
4
3
2
1
0
-1
-2
degree C
1982
1984
1986
1988
1990
1992
year

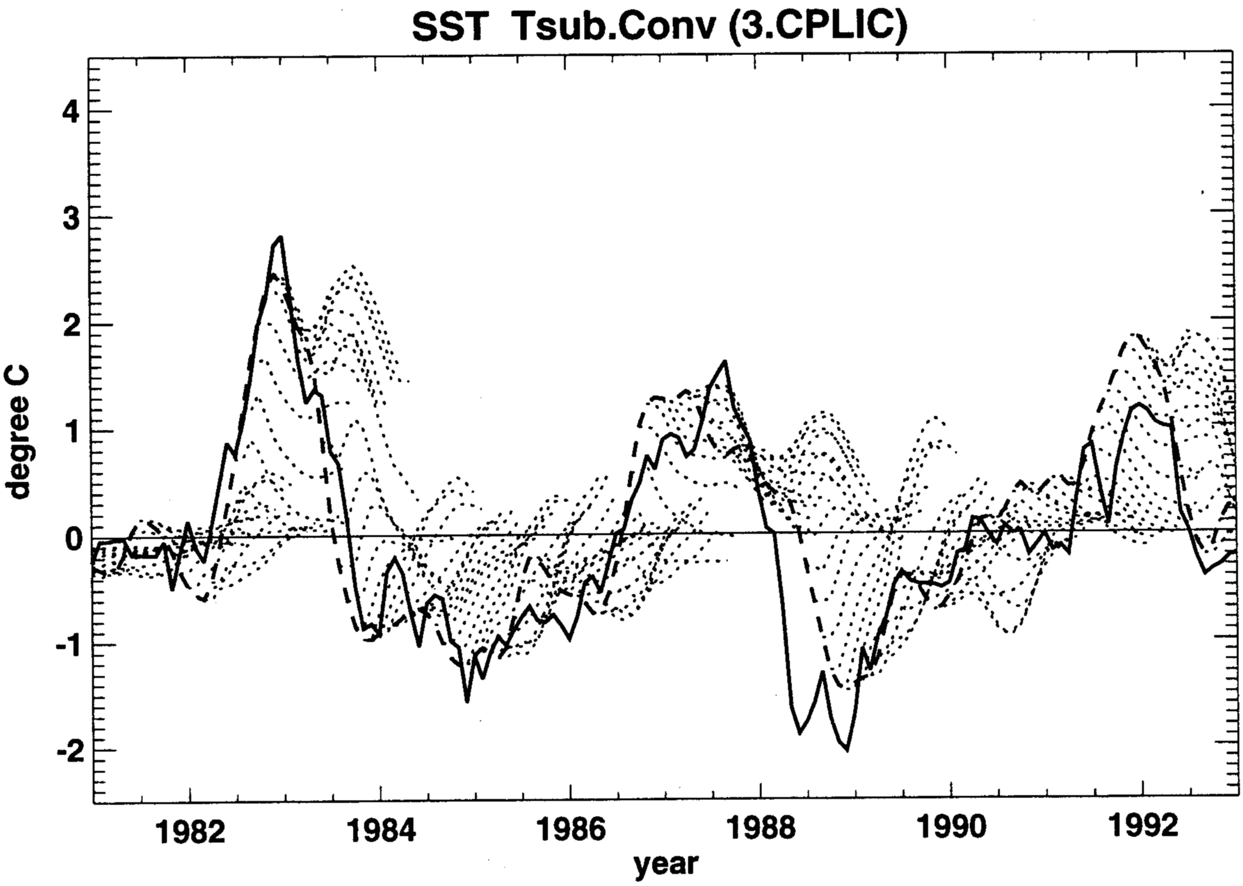
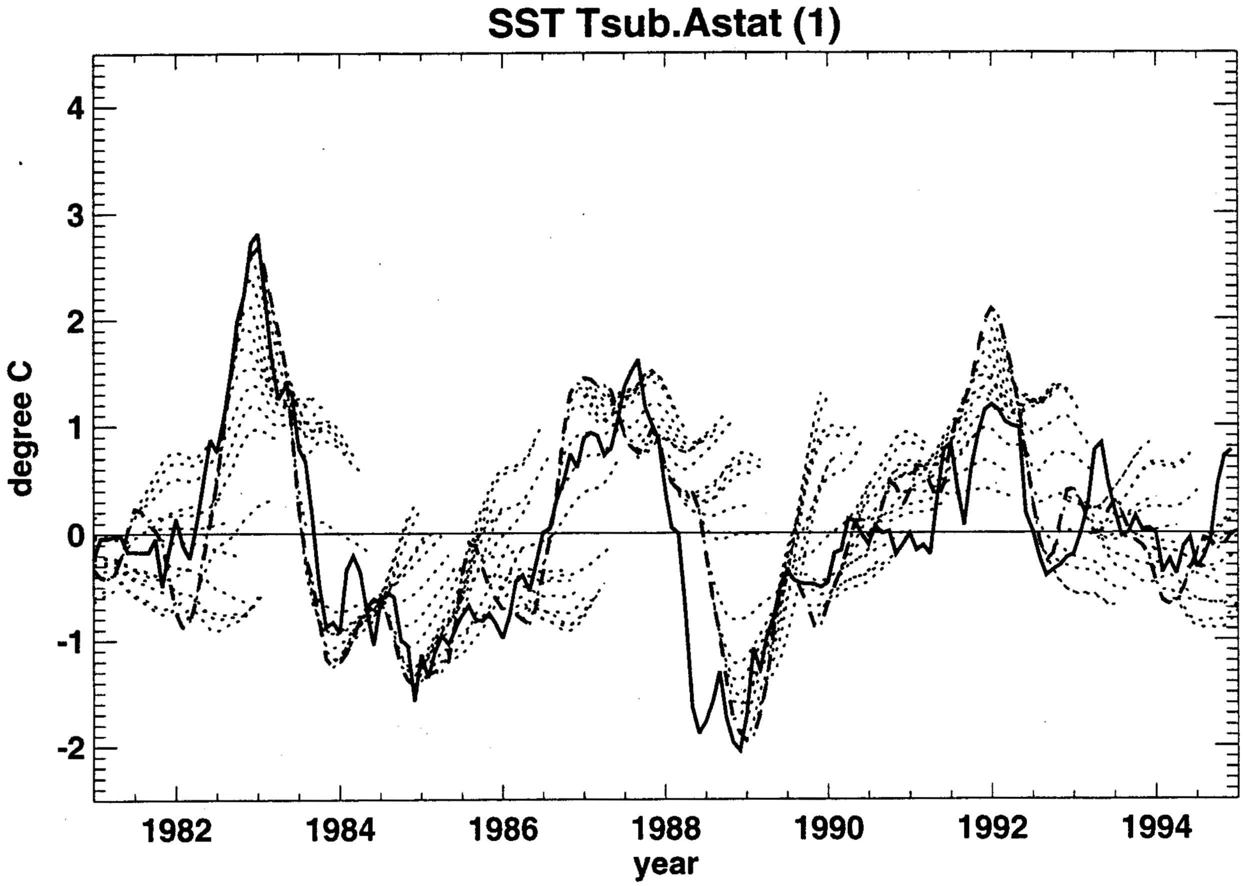

SST Tsub.Conv (3.CPLIC)
4
3
2
1
0
-1
-2
degree C
1982
1984
1986
1988
1990
1992
year

Fig.10de


SST Tsub.Astat (1)
4
3
2
1
0
-1
-2
degree C
1982
1984
1986
1988
1990
1992
1994
year

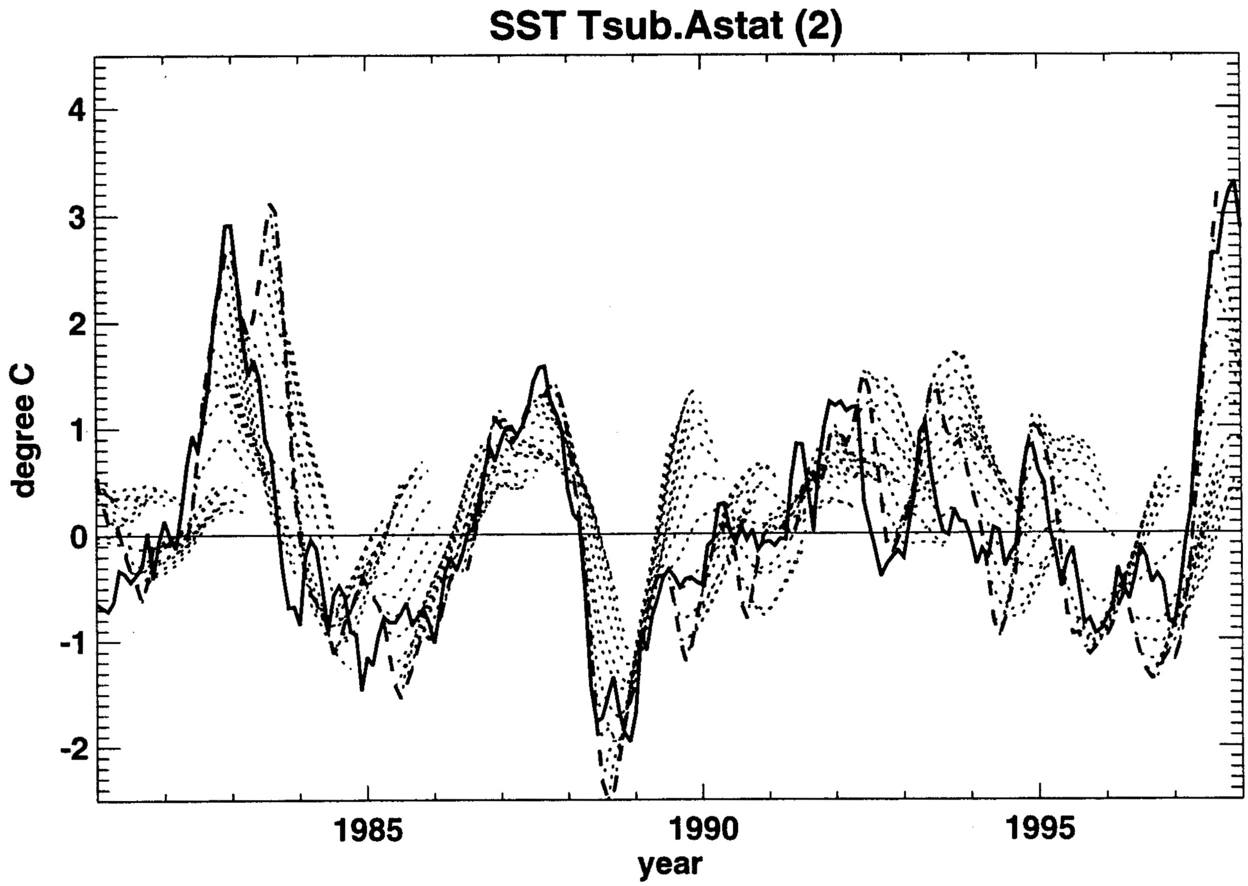

SST Tsub.Astat (2)
4
3
2
1
0
-1
-2
degree C
1985
1990
1995
year

Fig.11ab

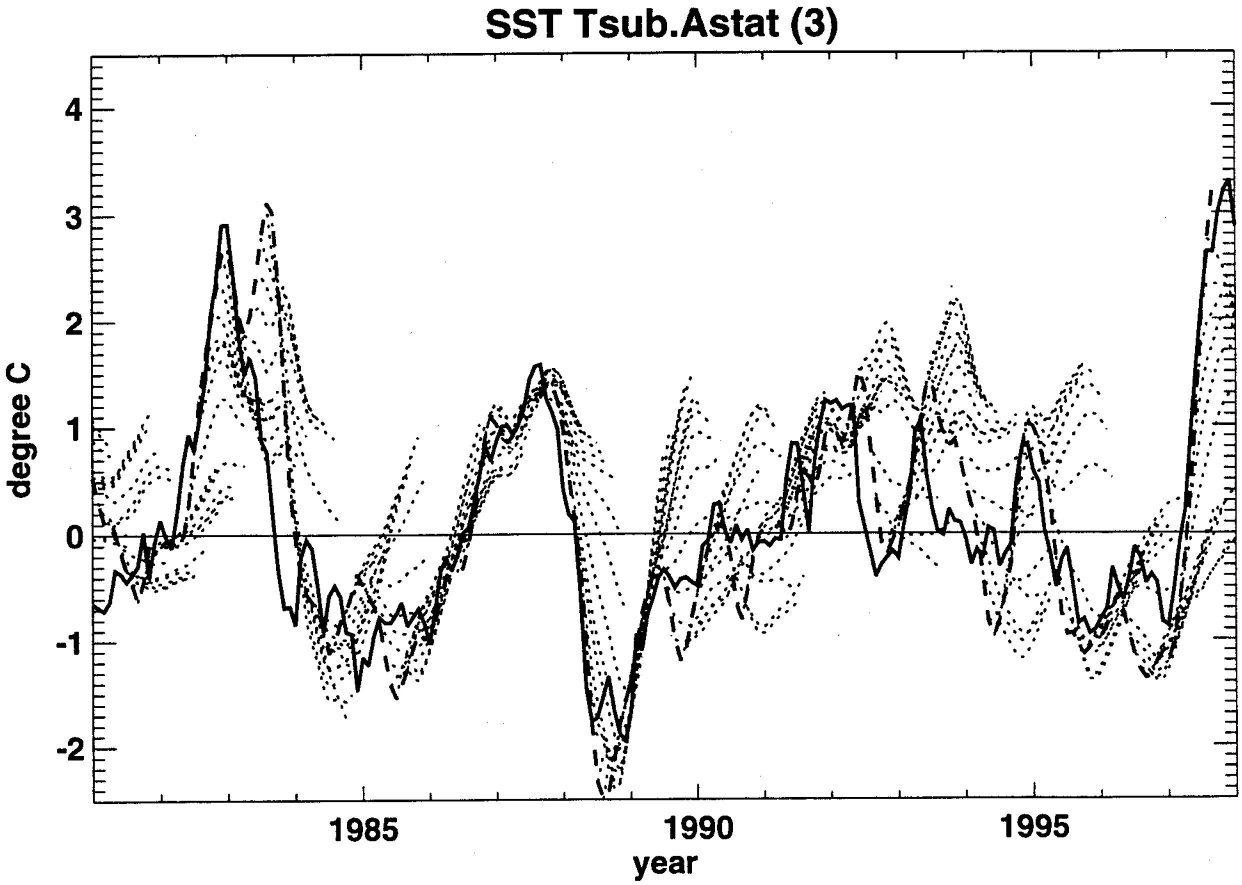


Fig.11c

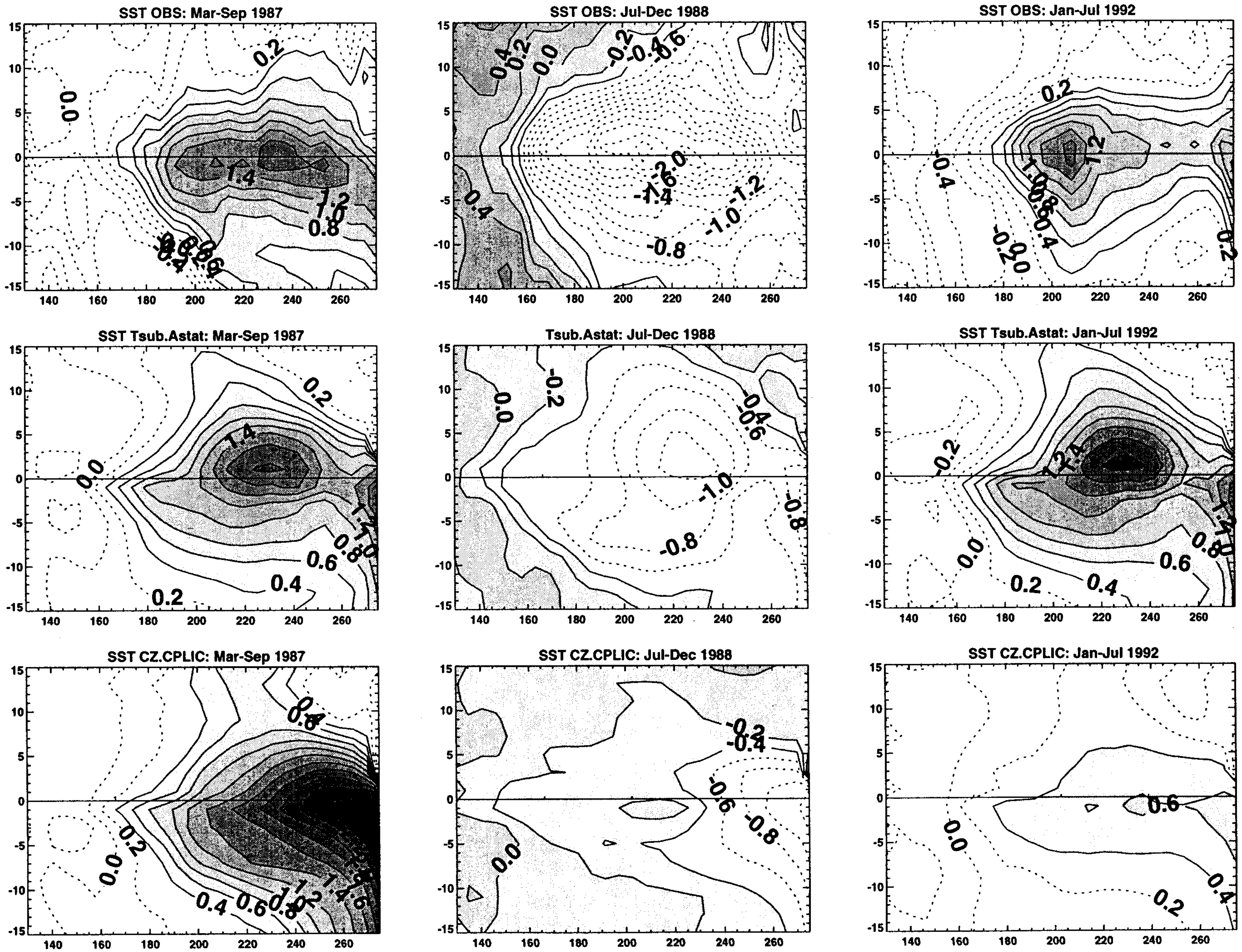
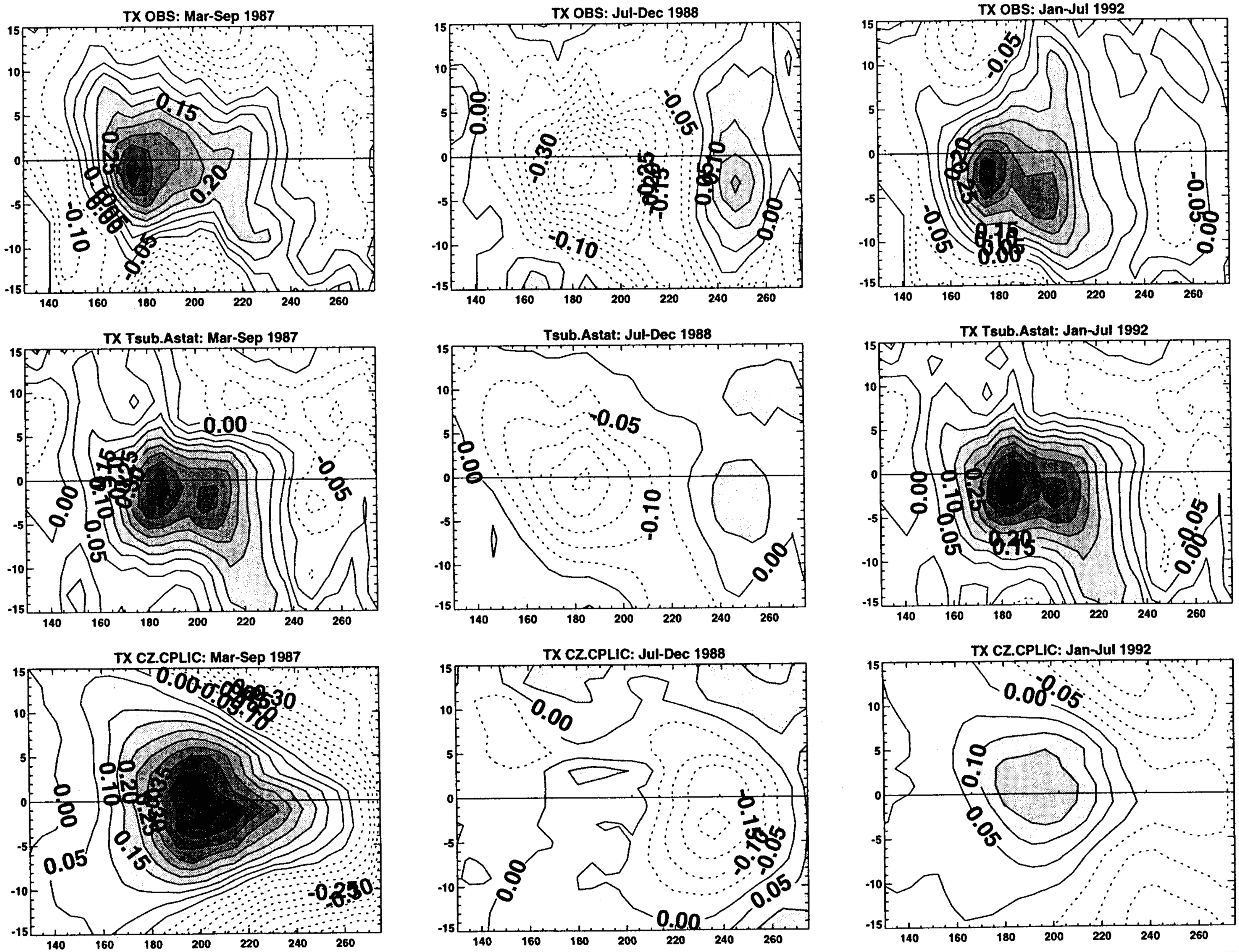
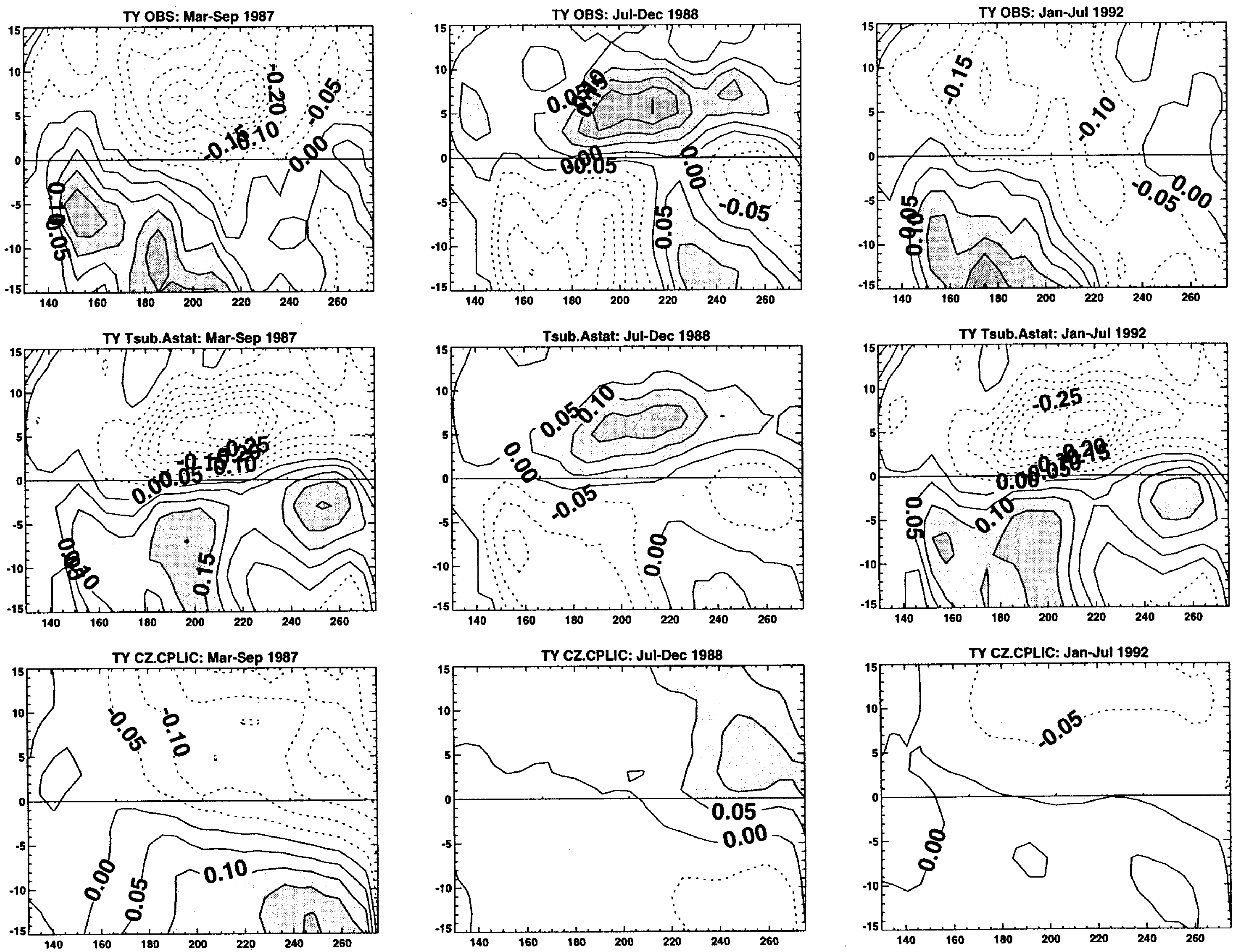

SST OBS: Mar-Sep 1987
SST OBS: Jul-Dec 1988
SST OBS: Jan-Jul 1992
SST Tsub.Astat: Mar-Sep 1987
Tsub.Astat: Jul-Dec 1988
SST Tsub.Astat: Jan-Jul 1992
SST CZ.CPLIC: Mar-Sep 1987
SST CZ.CPLIC: Jul-Dec 1988
SST CZ.CPLIC: Jan-Jul 1992

Fig.12



TX OBS: Mar-Sep 1987
TX OBS: Jul-Dec 1988
TX OBS: Jan-Jul 1992
TX Tsub.Astat: Mar-Sep 1987
Tsub.Astat: Jul-Dec 1988
TX Tsub.Astat: Jan-Jul 1992
TX CZ.CPLIC: Mar-Sep 1987
TX CZ.CPLIC: Jul-Dec 1988
TX CZ.CPLIC: Jan-Jul 1992

Fig.13



TY OBS: Mar-Sep 1987
TY OBS: Jul-Dec 1988
TY OBS: Jan-Jul 1992
TY Tsub.Astat: Mar-Sep 1987
Tsub.Astat: Jul-Dec 1988
TY Tsub.Astat: Jan-Jul 1992
TY CZ.CPLIC: Mar-Sep 1987
TY CZ.CPLIC: Jul-Dec 1988
TY CZ.CPLIC: Jan-Jul 1992

Fig.14

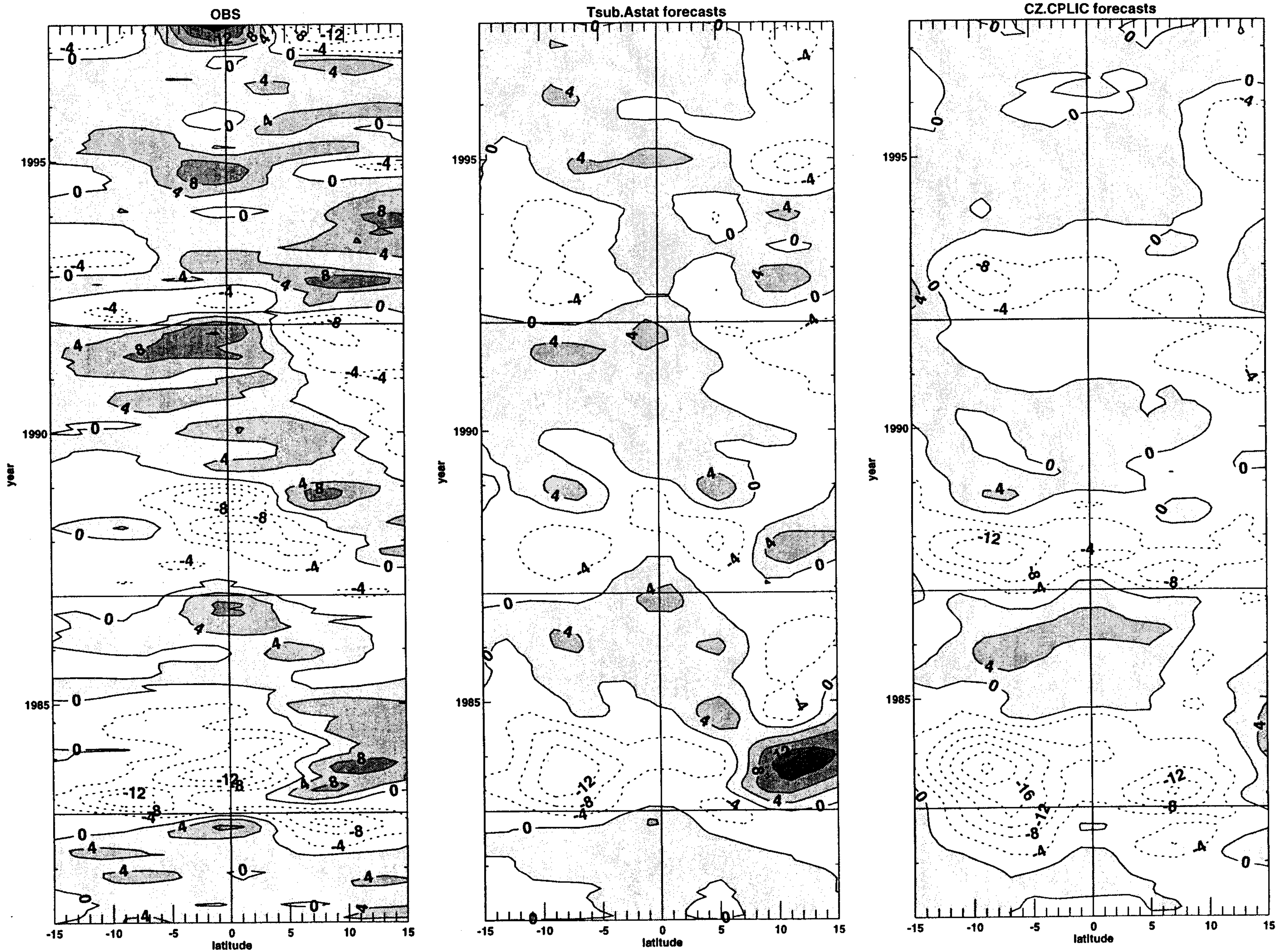

OBS
Tsub.Astat forecasts
CZ.CPLIC forecasts
year
latitude
1995
1990
1985
-15
-10
-5
0
5
10
15

Fig.15

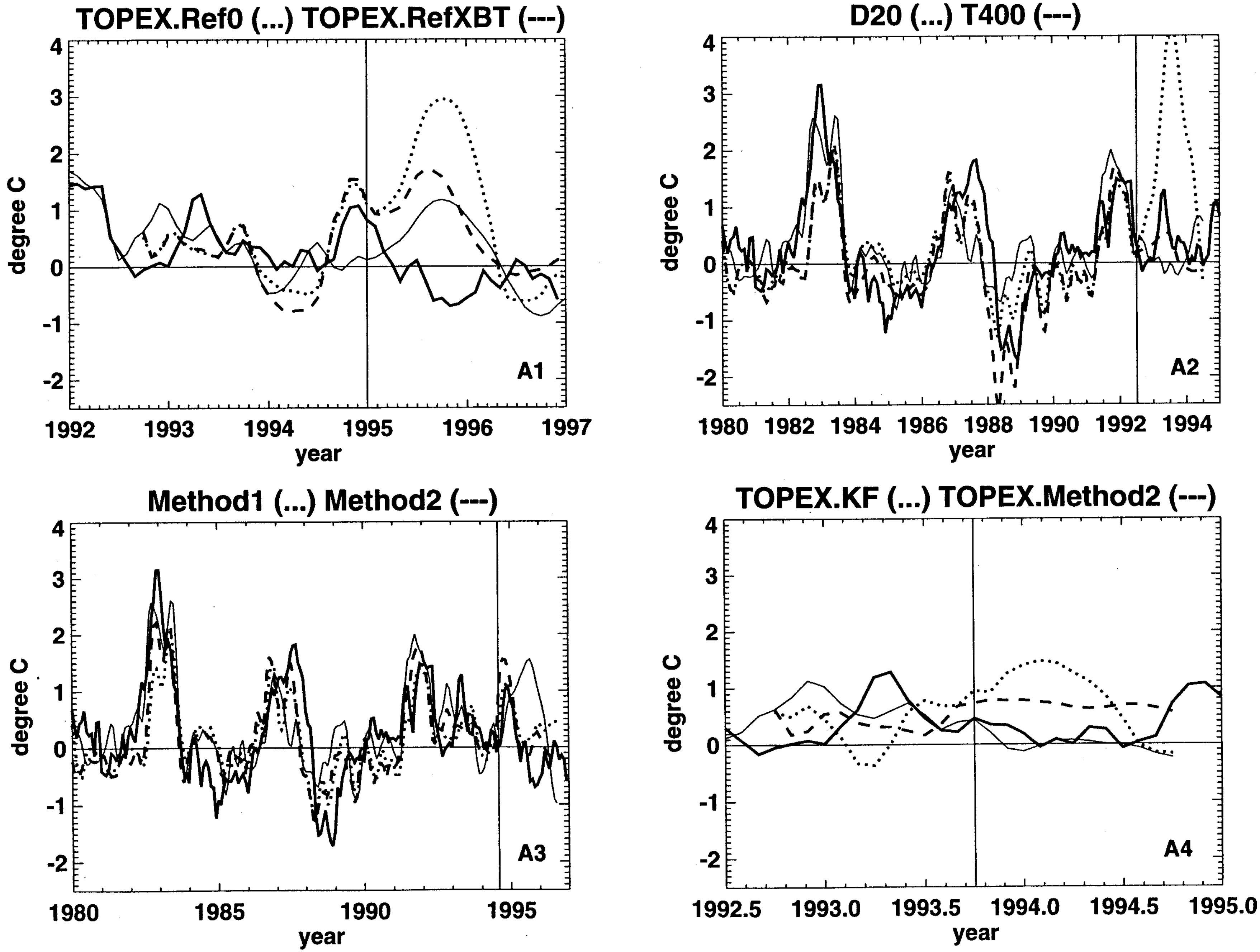
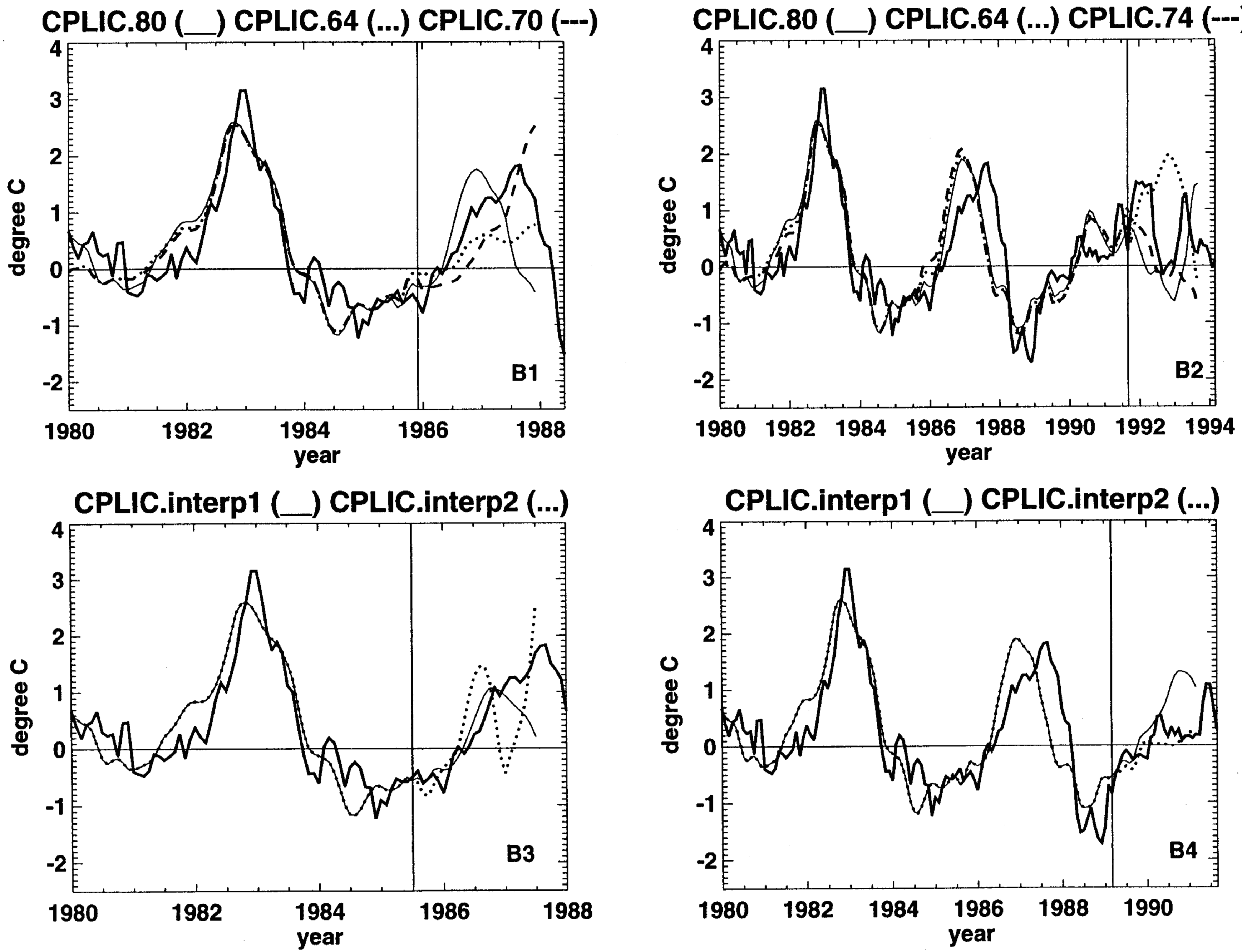

TOPEX.Ref0 (...) TOPEX.RefXBT (---)
D20 (...) T400 (---)
Method1 (...) Method2 (---)
TOPEX.KF (...) TOPEX.Method2 (---)
degree C
year
A1
A2
A3
A4

Fig.A


CPLIC.80 (__) CPLIC.64 (...) CPLIC.70 (---)
degree C
year
B1
CPLIC.80 (__) CPLIC.64 (...) CPLIC.74 (---)
degree C
year
B2
CPLIC.interp1 (__) CPLIC.interp2 (...)
degree C
year
B3
CPLIC.interp1 (__) CPLIC.interp2 (...)
degree C
year
B4